ERIC REPORT RESUME

ED 010 208 1-31-67 24 (REV)

STUDIES OF THE USE OF PROGRAMED INSTRUCTION IN THE CLASSROOM.

GLASER, ROBERT * AND OTHERS
TGZ65143 UNIVERSITY OF PITTSBURGH, LEARNING R AND D CTR., PA.

BR-5-0253-TR-1
TR-1
-MAY-66 OEC-2-10-057
EDRS PRICE MF-$0.27 HC-$7.76 194P.

*ELEMENTARY SCHOOL STUDENTS, *PROGRAMED INSTRUCTION, PROGRAMING, CLASSROOMS, CLASSROOM TECHNIQUES, *ARITHMETIC, *SPELLING, *EDUCATIONAL RESEARCH, CLASSROOM RESEARCH, MANAGEMENT, ACHIEVEMENT, EVALUATION TECHNIQUES, EVALUATION METHODS, TESTS, INDIVIDUAL INSTRUCTION, INDIVIDUALIZED PROGRAMS, RESEARCH AND DEVELOPMENT CENTERS, PITTSBURGH, PENNSYLVANIA

DURING THE SCHOOL YEAR OF 1963-64, THE LEARNING RESEARCH AND DEVELOPMENT CENTER (UNIVERSITY OF PITTSBURGH) AND THE BALDWIN-WHITEHALL PUBLIC SCHOOLS OF PITTSBURGH INITIATED AN EXPERIMENTAL PROJECT TO INVESTIGATE THE FEASIBILITY OF A SYSTEM OF INDIVIDUALIZED INSTRUCTION IN AN ENTIRE K-6 SCHOOL. PART ONE DESCRIBED THE INITIAL WORK ON THE USE OF PROGRAMED MATERIAL IN AN INTACT CLASSROOM. SPECIFIC STUDIES IN ARITHMETIC AND SPELLING AND THE INVESTIGATION OF STUDENT VARIABLES SUCH AS ATTENTION, ATTITUDE, AND APTITUDE WERE DESCRIBED. A SOCIOLOGICAL STUDY OF SOME OF THE UNANTICIPATED CONSEQUENCES OF EDUCATIONAL INNOVATION ON SUPERVISION, TEACHERS, AND SCHOOL ORGANIZATION WAS PRESENTED ALSO. PART TWO DESCRIBED EXPLORATORY STUDIES ON THE INDIVIDUALIZATION OF INSTRUCTION IN FLEXIBLE CLASSROOM CONTEXTS. IMPLICATIONS FOR THE USE OF PROGRAMED INSTRUCTION WERE--(1) PROCEDURES NEED TO BE DEVELOPED TO PERMIT EFFECTIVE MANAGEMENT OF INDIVIDUALIZED PROGRESS, (2) PROCEDURES SHOULD INCLUDE DETAILED DIAGNOSTIC ASSESSMENT OF STUDENT CAPABILITY, (3) RESEARCH AND DEVELOPMENT SHOULD BE DIRECTED TO METHODOLOGY FOR EVALUATION OF EDUCATIONAL ACHIEVEMENT, AND (4) ACHIEVEMENT TESTS SHOULD EMPHASIZE PERFORMANCE MASTERY, AND SUBSEQUENT USE AND REUSE OF KNOWLEDGE AND SKILLS. (JC)

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document [illegible] reproduced exactly as received from the person or organization [illegible] of view or opinions stated do not necessarily represent official Office of Education position or policy.

STUDIES OF THE USE OF PROGRAMMED INSTRUCTION IN THE CLASSROOM

Learning Research and Development Center
University of Pittsburgh

and

The Baldwin-Whitehall Public Schools
Baldwin Borough, Baldwin Township, Whitehall Borough
Pittsburgh, Pennsylvania

ED010208

Foreword by
W. Robert Paynter

Part One: Studies in the Intact Classroom

The Use of Programmed Materials in Elementary Arithmetic and Spelling
by Robert Glaser, James H. Reynolds, and Margaret G. Fullick

The Impact of Programmed Instruction on Selected Student Variables
by C. Mauritz Lindvall

Programmed Instruction: Some Unanticipated Consequences
by Richard O. Carlson

Part Two: Pilot Studies of the Individualized Classroom

The Use of Programmed Materials to Individualize and Extend Instruction
by James R. Hawker and Robert Glaser

ED010208

May, 1966

The research reported herein was supported through the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare, Contract Numbers OE 2-10-057 and OE 4-10-158.

FOREWORD

The Baldwin-Whitehall Public Schools of suburban Pittsburgh have been cooperating with institutions of education for many years. Our work with universities and colleges arises from our fundamental belief that innovations in education can most fruitfully develop from this kind of association. It is the important role of school boards and school administrators to become alert to new educational developments and also to sensitize researchers to a school system's requirements for research and development.

During the 1961-62 school year, the decision was made to study the effectiveness and impact of the new programmed instruction in our schools. A series of studies was instituted at all levels from elementary school through high school. However, only the work which we think will be of the most wide-spread interest is described in this report. In general, we learned some things about this new educational tool. Specifically, we found that the principles of programmed instruction were useful concepts, but that particular programmed instructional materials have to be evaluated on the basis of their own merit, which is in fact a principle of the programmed instructional procedure itself.

Most significantly, the use of programmed instructional materials with their emphasis on individualized learning awakened within us the educator's long-time goal of producing innovations in education that would permit a system which would be highly adaptable to individual student needs. We are pursuing this with the Learning Research and Development Center at the present time. As indicated in the Preface, the studies reported here served as a prelude for our present projects with the Center.

In my role as Superintendent of the Baldwin-Whitehall School System, I must acknowledge those individuals who particularly encourage a climate

of innovation and who are particularly helpful in carrying out our continuing quest for quality in education. These people include Dr. J. Steele Gow, Jr., Dr. Warren D. Shepler, Dr. J. Ernest Harrison, Mr. Arthur D. Jeffries, and Mr. Howard F. Phillips.

Dr. W. Robert Paynter

Pittsburgh, Pennsylvania
May, 1966

PREFACE

During the school year of 1963-64, the Learning Research and Development Center and the Baldwin-Whitehall Public Schools of suburban Pittsburgh initiated an experimental project to investigate the feasibility of a system of individualized instruction in an entire K-6 school. This came about as a result of a series of prior exploratory studies, begun in 1961-62, designed to test preliminary notions on a smaller scale and in single classrooms. The work started with the use of programmed instruction in an intact classroom, the intact classroom being defined as a classroom unit in which the teaching practices were oriented around the conventional grade-by-grade progression of learning.

As work proceeded, it soon became apparent that the significant individualization feature of programmed instruction could not be manifested unless the intact classroom changed its organization to permit a more flexible progression. As a result, a second set of studies was instituted to use programmed instruction and other materials in a more flexible context. Out of this experience grew the Individually Prescribed Instruction Project currently in progress, in which various combinations of instructional materials, including programmed materials, special workbook and test procedures, and teacher practices, are being used for the purpose of adapting to individual student requirements.

Part One of this report describes the initial work on the use of programmed material in the intact classroom: Chapter 1 describes specific studies in arithmetic and spelling; Chapter 2 describes the investigation of student variables such as attention, attitudes, and aptitude; and Chapter 3 describes a sociologist's study of some of the unanticipated consequences of an educational innovation on supervision, the teacher, and school organization. Part Two, which consists of Chapter 4, describes exploratory studies in the individualization of instruction in flexible classroom contexts.

Primary acknowledgement for the completion of these studies goes to the Baldwin-Whitehall teaching staff and the principals of the various schools involved. Over the course of the studies, all of which are not reported here, more than fifty teachers and two thousand students participated.

Particular acknowledgement must go to the senior administrators of the Baldwin-Whitehall Public Schools: W. Robert Paynter, Superintendent, Warren D. Shepler, Assistant Superintendent in Charge of Instruction,* and J. Ernest Harrison, Director of Curriculum and Research. The foresight of these men has been a constant source of encouragement for the work of the Center.**

Mr. Jack R. Fisher coordinated the conduct of the studies and data-collection procedures. He also supervised the writing of supplementary teacher materials and prepared preliminary drafts for the writing of this report. During 1961-62, Mr. Theodore A. Harakas monitored the conduct of the studies and contributed to analysis of the data.

***Now Assistant Superintendent of the Harrisburg Public Schools in Harrisburg, Pennsylvania.**

****Now Assistant Superintendent in Charge of Instruction of the Baldwin-Whitehall Public Schools.**

Table of Contents

FOREWORD by W. Robert Paynter i

PREFACE iii

GENERAL INTRODUCTION AND IMPLICATIONS 1

PART I: STUDIES IN THE INTACT CLASSROOM

CHAPTER I

The Use of Programmed Materials in Elementary Arithmetic and Spelling

Robert Glaser, James H. Reynolds, and Margaret G. Fullick

INTRODUCTION 7

The Nature of the School System 8

The Questions Asked 9

Control Aspects 9

The Processing Data 14

ARITHMETIC TOPICS IN GRADE I 15

Exploratory Study: The Classroom Use of Simple Teaching Machines with Young Children 22

Study 1-A: The Effect of Varying Teacher-Program Combinations Upon Learning 32

Study 1-B: Effects of Prefamiliarization, Distribution of Instruction, and Post-Learning Practice Upon Programmed Instruction in Telling Time 38

Study 1-C: Comparisons of Program and Non-Program Groups at the End of Grade 1 54

ARITHMETIC AND SPELLING IN GRADE 4 60

Overview 60

Study 4-A: The Effect of Intelligence Upon Programmed Learning Achievement 65

Study 4-B: The Use of Programmed Instruction for Review and Acceleration of Arithmetic Learning 70

Study 4-C: Effects of Classroom Surroundings Upon Programmed Learning in Spelling 79
IMPLICATIONS 87
REFERENCES 88

CHAPTER II

The Impact of Programmed Instruction on Selected Student Variables

C. Mauritz Lindvall

INTRODUCTION 91
STUDIES OF PUPIL ATTENTION AND PUPIL ATTITUDE 93
  Studies of Pupil Inattentiveness 93
    First-Grade Students Using a Time-Telling Program . . 94
    Fourth-Grade Students Using an Arithmetic Program . . 95
    Fourth-Grade Students Using a Spelling Program . . . 97
  Studies of Pupil Attitude 100
    Changes in Attitude Toward the Use of Programmed Materials 101
    The Correlation of Attitude with Achievement and Intelligence 103
  Changes in Attitude Toward the Subject Being Studied 104
STUDIES OF THE CORRELATION OF ACHIEVEMENT AND APTITUDE 112
  The Correlation of Aptitude and Achievement for Fourth-Grade Students Using Spelling and Multiplication and Division Programs 113
  The Correlation of Intelligence and Achievement for Seventh-Grade General Science Students 114
SUMMARY 118
REFERENCES 119

CHAPTER III

Programmed Instruction: Some Unanticipated Consequences

Richard O. Carlson

INTRODUCTION 121
CLASSROOM OBSERVATION AND TEACHER SUPERVISION 123
RATES OF INDIVIDUAL ACHIEVEMENT 126
THE TEACHER'S NEED TO PERFORM 129
SUMMARY 131

PART II: PILOT STUDIES OF THE INDIVIDUALIZED CLASSROOM

CHAPTER IV

The Use of Programmed Materials to Individualize and Extend Instruction

INTRODUCTION 135
INDIVIDUALIZED INSTRUCTION IN SPELLING 137
INDIVIDUALIZED INSTRUCTION IN ARITHMETIC 150
INDIVIDUALIZED USE OF PROGRAMMED INSTRUCTION AS A REVIEW TOOL . . 173
SUMMARY AND CONCLUSIONS 179
REFERENCES 182

GENERAL INTRODUCTION AND IMPLICATIONS

The most useful introduction to this report would seem to be a description of findings and observations which have broad implications for future studies on instructional procedures in the classroom. Details, specific data and elaboration of conclusions are to be found in the body of the text. The motivation behind the series of studies carried out was to examine student instruction and achievement in basic subjects taught largely by programmed instruction in the elementary school and to suggest what this examination might mean for the improvement of instructional practice and for the improvement of research methodology on classroom learning. The work presented in this report generally suggests the following with respect to the student, the teacher, and research methodology.

Student Learning and Achievement

1. The student can obtain more information and knowledge than he presently obtains. When the time of a lesson period is kept constant and additional individualized instruction is added, extended learning in advanced topics can be provided. The time for this extended learning takes away from extra practice in regular work, but test results show that the advanced work is learned without detriment to the usual grade level attainment. Also, with carefully programmed material, practice and review carried out at the beginning of a school year can be accomplished in a shorter period of time than usual.

2. Generalized measures of intelligence are not necessarily related to student's tested achievement. The relationship of intelligence to achievement depends on the instructional method employed. With effective individualized instructional procedures, aptitude and intelligence measures may be more related to rate of learning than measured achievement in a particular unit of learning.

3. The variability of progress and the differential accomplishment of students indicate the urgency for the development of systems of instruction which adapt to the individuality of the student. Students differ over

a very wide range with respect to the amount of time they require to attain mastery of a learning unit. Furthermore, when students are pretested on the contents of a lesson prior to instruction, a significant number of students show knowledge of what they are about to learn or the absence of adequate prerequisite knowledge and skills.

4. Self-instructional activities using programmed materials apparently maintain student attentiveness. Observations of students during courses of programmed instruction indicate no increase in boredom or inattention.

Teachers and Supervisors

5. The use of programmed materials makes difficult the usual procedures that supervisors carry out to observe classroom instruction. The supervisor is accustomed to observing daily or weekly lesson plans and observing the teacher working with the class as a group. With the use of self-study materials, much of the lesson plan is built into the programmed materials, and teacher activity with an entire class of students is decreased. If procedures for the individualization of instruction are implemented in a school, new thought must be given to the way in which supervisors evaluate teaching effectiveness.

6. The usual structure of a school with its intact classroom makes the management of individualized instruction very difficult for the teacher. When programmed self-instructional materials are used which permit the students to spread themselves out over a wide continuum of achievement, the teacher must develop techniques to control this range. If the school organization does not change to meet the range of student achievement, it is necessary for teachers to employ techniques which minimize the extent of individualization so that they can perform their job as classroom teachers effectively.

7. Most current pedagogical procedures taught in schools of education instill teachers with a need to hold the attention of the student and to constantly mediate between the student and the subject matter. In contrast, self-instructional procedures preclude certain teacher activities

of this kind and give the teacher the feeling that he is really not teaching. As a consequence, teachers may frequently modify the use of self-instructional materials to permit "real teaching" to occur. The new role that the teacher must learn in individualized instruction is one which emphasizes more detailed evaluation of student performance. This requires more detailed prescription and guidance based on this evaluation than is the case in standard teaching. This role can be a highly professional and demanding part of the teacher's time and requires appropriate teacher preparation.

Research Methodology and Evaluation

8. When studies of individualized instruction are compared with the hard-to-define variable called "standard instruction," the results obtained in this report and in previous studies in the literature show a general pattern: self-instructional programs result in student achievement which is about the same or better than standard instruction; few spectacular differences have been shown here, and indeed in the history of recent innovation in classroom instruction. A number of hypotheses can be suggested to explain this; one hypothesis which has been suggested in the course of this work is the inadequacy of measuring instruments for subject-matter achievement.

To some extent it seems that test construction procedures are weighted against the display of the effects of different experimental instructional treatments. In particular, published tests of subject-matter achievement are constructed to have specified levels of item difficulty. These levels of item difficulty permit the tests to differentiate among the students in order to provide a discriminating measuring instrument; but what artifacts are built into the test in order to provide levels of difficulty? Suppose a child might have mastered the multiplication of three-place and four-place numbers; when an appropriate sample of items is employed to test him, the student should get most of them right or most of them wrong if he has mastered the appropriate skill. It is conceivable that procedures for building difficulty levels into tests oppose the clear-cut measurement of subject-matter mastery and add an additional dimension to the test. In order to provide differential difficulty, features are brought into the test which require incidental learning on the part of the student. This incidental learning may be relevant or irrelevant to the subject-matter objectives at hand, but this incidental learning may not be irrelevant to general intelligence. It can be hypothesized that selecting item difficulty levels for a test in order to make it more discriminating makes it more a measure of IQ and thus lessens the relevance of the test to the effect of the instructional treatment which teaches a particular skill. Once multiplication of two-place numbers is taught effectively, what kind of difficulty levels are appropriate for measuring performance on this task: Should this measurement include the ability to transfer this learned skill to work problems and to new situations? If so, should this ability be taught or should it emerge as a function of general intelligence? A suspicion growing out of the sequence of studies in this report is the necessity for careful analysis of achievement measurement procedures which are relevant to the experimental comparisons being undertaken.

9. End-of-course achievement examinations may be only one aspect, and a minor one, in the measurement of student achievement. "Learning" of the fundamentals of the subject-matter areas may be only superficially tested by end-of-course performance. The deeper effects of instruction should show up in long-term measures of the recall of knowledge in appropriate situations or in the ability to relearn a skill prior to its reuse.

10. In general, the implications of the use of programmed instruction in the classroom are the following: (a) Procedures need to be developed to permit the effective management of individualized progress; (b) Individualization procedures will include detailed diagnostic assessment of student capability of the basis of which instructional plans will be developed for him; (c) Research and development effort must be devoted to methodology for the evaluation of educational achievement; (d) Tests of student achievement should emphasize performance mastery and the subsequent use and reuse of knowledge and skills. Normative comparisons among students which do not specify the extent of mastery of subject-matter objectives may be of minimal use for the improvement of teaching procedures.

PART ONE:

STUDIES IN THE INTACT CLASSROOM

PRECEDING PAGE BLANK-NOT FILMED

CHAPTER I

THE USE OF PROGRAMMED MATERIALS IN ELEMENTARY ARITHMETIC AND SPELLING

Robert Glaser, James H. Reynolds, Margaret G. Fullick

Introduction

When the use of programmed instructional materials is studied in a school system, two different approaches to implementation can be considered. One approach considers the concept of programmed instruction as a means for individualizing the instructional process. The other approach conceives of the program as a reproducible event that can be employed as a standard instructional sequence around which certain variations in classroom procedure can be effected.

The first notion, that of individualization, is a primary assumption behind the development of programmed instructional procedures. Ideally, programmed instruction is a means whereby the student can be provided with instruction on the basis of his particular requirements. A tutorial process is the analogy of the individualization process. The efficient tutor determines in detail the knowledge and skill that the student has prior to instruction; he then begins instruction assuming only the competences that the student has shown. The instructional procedure is adjusted for the student by the tutor according to the rate at which the student learns, the kinds of forward steps the student can take, and the kinds of experiences which the student finds rewarding and motivating for effective attainment of subject matter skills.

Current use of programmed instruction has far from attained the ideal of the individualization of instruction. At the present time, however, programmed materials and the concepts underlying them represent a step toward the provision of an individualized instructional environment for each student: they can permit the student to learn at his own rate and present him with the freedom to move ahead or catch up depending upon his mastery of the subject matter; they can also permit different students to study different subject matters at the same time in the same classroom. When programs are used with such individualization in mind, they obviously necessitate restructuring of the intact classroom unit because different students in the class will require different instructional conditions. Such reorganization is considered desirable by many school administrators but is a major problem for a school system where the unit of organization is intact class groupings and yearly grade-by-grade advancement.

Within the intact classroom structure, experimentation with programmed materials takes the form of treating the program as a standard instructional technique which can be used in various ways in the classroom by manipulating certain aspects of the classroom instructional procedure. In this way, the achievement of the class and the manipulation of classroom teaching procedures become, respectively, the dependent and independent variables for study.

For the most part, the introduction of programmed instructional materials in the Baldwin-Whitehall Schools took place in intact classroom groups. The programs used were those commercially available at certain grade levels, and the variations studied were essentially manipulations of classroom procedure. The studies reported here took place in the academic year 1962-63 and with the exception of those in grade 9, represent attempts to examine the use of programmed instructional materials within existing classroom structures.

The Nature of the School System

The studies reported here have been carried out in the Baldwin-Whitehall Public Schools, situated in a suburban residential area contiguous to the City of Pittsburgh. The population of the area represents a cross-section of the metropolitan Pittsburgh area, ranging from skilled mill and industrial workers to executive and professional types. At the time of the studies the physical facilities of the school system consisted of one high school, two junior high schools, and 12 elementary schools (kindergarten through sixth grade). The total student enrollment during the 1962-63 term was approximately 8,000, with a classroom teaching staff of approximately 375.

The Questions Asked

The questions which were asked about the use of programs arose from primarily two sources: (1) variables studied in the psychologist's learning laboratory that suggested a procedure for improving instructional effectiveness, e.g., the distribution of practice, and (2) problems arising from general teaching practices and educational requirements, e.g., the necessity for providing extended opportunities for learning (acceleration). Sometimes both of these sources provide the background for a particular experiment.

Studies were designed to investigate the following kinds of questions:

Grade 1. Can simple teaching machines be used in the classroom with young children beginning the first grade? What is the relative effectiveness of different teacher-program arrangements upon learning? What is the relative effectiveness of varying the distribution of daily work with the program? What is the effect of prefamiliarization and post-learning practice in the achievement resulting from programmed instruction? At the end of the school year, do classes using programs in arithmetic topics compare with classes not using such programs?

Grade 4. What is the relationship between intelligence and use of programmed instruction under certain conditions? How effective is programmed instruction for the review and acceleration of learning? What are the effects of classroom surroundings upon learning from a program?

Grade 7. What is the effect of various combinations of programmed instruction and "enrichment" activity? Does prefamiliarization and an overview of material to be learned improve the effectiveness of programmed instruction?

Grade 9. How effective is programmed instruction in providing the opportunity for learning additional subject material?

Control Aspects

When specific studies are set up in an on-going school situation to answer these questions, a number of variables must be considered which can influence the data obtained. The influence of these variables must be considered in interpreting the results of the studies or must be controlled in some way. The following aspects were of concern in the studies reported.

The Quality of the Programmed Instructional Materials. With the exception of one program constructed at the University of Pittsburgh, the programs employed were commercially available from reputable program publishers. These publishers provided some evidence that the programs were constructed according to good program development practices and were effective instructional instruments. This evidence was of informal nature, since most program publishers at the present time do not provide manuals giving detailed data on program use and validity. It is anticipated that manuals similar to those accompanying nationally standardized tests, containing the validation data obtained during the course of program development, will be made available in the future. Furthermore, standard criterion which publishers can follow in the development of a program manual are being developed by national committees (Joint Committee on Programmed Instruction and Teaching Machines, 1963).

The degree of effectiveness of the various programs, especially for the population of students participating in the separate studies, was not specifically known prior to use, and the efficiency and effectiveness with which they taught varied. The extent to which the effectiveness of a program interacted with the particular study being carried out is difficult to assess, and the differences in this variable of instructional quality was controlled only to the extent that some impression was available about initial program construction and subsequent development and use.

The Subject Matter. The subject matter taught by programs in the present studies was selected on the basis of (a) availability for the particular grade levels involved and (b) subject matter requirements in terms of student need, school requirements, and student-teacher-community acceptance as determined by the Baldwin-Whitehall school administrators. Of the eight commercial programs used, six pertained to arithmetic or mathematics. This reflects the fact that a preponderance of programs available at this time were on mathematics topics and that this is a topic readily introduced into school systems in program form; the two other subject matters were spelling and general science. The extent to which the greater number of programs in mathematics influenced the results of the studies carried out is again difficult to assess.

Teacher Characteristics. In all of the studies, the teacher participated, to a greater or lesser degree, in instruction in the subject matter area involved in the program. As a result, teacher characteristics could influence the data obtained. However, since intact classes were used, teacher characteristics were controlled to the extent that at least two different teachers were involved in each of the experimental conditions compared. This limited control had the effect of preventing any one experimental condition from dependence upon a single teacher. In addition, all teachers participating in the various studies were chosen on the basis of a positive (or at least a non-negative) interest in trying out programmed instructional materials. Prior to classroom introduction, teachers participated in the development of the particular procedures to be used; one teacher for each study at each grade level prepared a manual for all teachers involved in that particular study. This manual consisted of a day-by-day plan of the specific classroom activities that would be carried out for the subject matter being studied. The exact manner in which the program was to be used was specified, and teacher materials for non-program instruction were elaborated in detail. In this way, some standardization of the procedures being studied was accomplished. In addition, a research coordinator checked with the teachers several times each week in the course of a particular study to insure that procedures were being carried out as planned.

Student Ability. Past experience in the Baldwin-Whitehall Schools has indicated that differences among classes in intelligence levels and previous subject matter achievement influence learning from programmed instruction. This is so despite the often-quoted claim that the individualization offered by programmed instruction will reduce the relationship between student attainment and measured intelligence to near zero.

There are a number of factors involved in assessing this statement which have been discussed elsewhere (Glaser, 1963). The fact is that with the programs used and the manner in which they were used in the studies reported here, such a relationship between aptitude and achievement does exist. As a result it has been necessary in the separate studies reported to control the classes compared on the basis of average intelligence and achievement levels, and comparisons to assess the effects of the independent variables have required careful matching of class means in order to draw appropriate conclusions.

Testing Procedures. Crucial to the assessment of experimental effects are the measures employed to test the dependent variable, student achievement. In assessing the outcome of programmed instruction various measures can be used, each of which has particular characteristics. Three main types of measures can be distinguished, namely, program tests, teacher-made tests, and nationally-standardized tests. Program tests are achievement tests which accompany the program and which the program publisher considers an adequate sample of student performance of the objectives taught by the instructional sequence. Teacher-made tests are those tests developed in cooperation with the classroom teacher and consist of items representative of the educational objectives of classroom instruction. Nationally-standardized tests are those commercially available achievement tests used by schools to assess their instruction and compare themselves with national norms. All three types of tests were employed in the various experimental studies reported. Where the program test was not considered an adequate test of overall classroom objectives or of the program itself, it was supplemented by a teacher test or a nationally-standardized test. When a nationally-standardized test was used, agreement was obtained from the teacher and school administrators that this test was a satisfactory measure of their own course objectives.

Ceiling Effects. If in assessing experimental variations in the classroom, a definitive test is established to indicate mastery of the course objectives, then the objectives of instruction are to teach so that students attain such mastery. This means that in successful instruction many students will obtain perfect scores and the distribution for a class will be skewed in accordance with a ceiling imposed by perfect test performance. If two different instructional treatments are given to two different groups and both groups show many students with near perfect test scores, there is then the question of distinguishing which treatment represents the more effective instruction. Factors other than student achievement, such as time taken to attain mastery, etc., must be considered. If achievement is the only measure of concern, however, then the percentage of students obtaining a perfect score, the average level of mastery or the gain in mastery from pre- to posttesting can be used. A question might always remain, however, with respect to how much more knowledge would have been exhibited by students if the test did not have a mastery ceiling. For example, if the objective of a course of instruction is to teach students addition and subtraction with single-digit numbers, a mastery test would measure just that skill, addition and subtraction with single-digit numbers. However, it is justifiable to ask to what extent students can extrapolate and transfer their knowledge to two- and three-place numbers. The tests employed in the studies reported here are, for the most part, tests with mastery ceilings and were used to assess the attainment of specific mastery objectives. Sometimes tests of more general objectives were employed which did not display ceiling effects. These were usually nationally-standardized tests constructed so as to give a wide distribution of scores. As will be seen in the studies reported, assessments of the various dependent variables are considered in the light of the characteristics of the achievement tests employed.

Extrapolation of Laboratory Findings. As has been indicated, a number of the studies reported involved variables suggested by laboratory experiments. In general, the direct extrapolation of a laboratory variable to actual instructional practice in intact classes runs many risks. One is that in group experiments in the laboratory the differences between experimental and control groups are often obtained under stringently controlled laboratory conditions and it can be expected that an effect of small magnitude under such control conditions will be attenuated in the conditions of the practical classroom. For the most fruitful interaction between the laboratory and instructional procedures in the classroom to take place, a research and development sequence is required which passes through fundamental laboratory research, through development, through design and proving and field tryout (Gilbert, 1962; Glaser, 1964).

The Processing of Data

All data obtained on both pre- and posttests were punched on cards and appropriate statistics obtained on the University of Pittsburgh's IBM 7070 computer. In working with intact classroom groups, individual cases are lost for a variety of reasons such as school transfer, absenteeism, etc. In addition, cases were eliminated in the attempt to match the classes compared, and the procedure whereby cases in the samples reported in this chapter were selected need to be specified here. For the experimental classes, no students were eliminated in any study because they obtained an exceptionally high or low score on an independent variable. Students were eliminated, however, if matching data on them were not available, i.e., IQ data and Stanford test achievement scores. Students were also eliminated if they missed more than one of the major subject matter posttests. Students who were repeating the grade, of which there were relatively few, were not included in the study, and students who left the school district during the year were not included. On the basis of these criteria, approximately 4% of the students in a class were eliminated for any of the above reasons.

Where data were missing on any one student, for example, only one posttest, the mean of the call for the variable involved was used to fill in the missing data for computer analysis. This procedure was used for less than 1% of the data obtained.

In the control classes (receiving conventional instruction) no students in the first and fourth grade were eliminated; in the seventh and ninth grade study, students quite low in IQ, below 90, were eliminated for matching purposes. In the seventh grade, for matching purposes, students were eliminated if no Stanford Achievement scores were available at the beginning of the year; these scores were not available because some students entered the seventh grade from a six-year parochial school.

Arithmetic Topics in Grade 1

Overview

Three main studies were carried out at the first grade level, using programs to present arithmetic topics that constituted a significant portion of the first grade curriculum. These topics included learning the addition and subtraction facts and learning to tell time. In addition to these two topics, a program which taught students to recognize and write the numbers from 1 to 10 was used in an exploratory way. Selection of topics was based primarily on the availability of program appropriate for the first grade. Two of the programs were available commercially, and the third (time-telling) was an unpublished experimental program undergoing development at the University of Pittsburgh.

Exploratory Study

The exploratory study examined the feasibility of presenting programmed instruction in a simple teaching machine on a group basis to six-year-old children. The child entering first grade cannot read, is not yet adept at following complex directions, and may not yet have developed the motor and perceptual skills required for such tasks as using a pencil, turning machine knobs, loading a machine with programmed materials, matching responses to feedback stimuli, etc. The objective of the exploratory study was to determine the effect upon learning when children in the first grade were given programmed instruction in a format requiring the use of a simple hand-operated machine.

Teacher-Program Coordination

The first more formal study (1A)[1] was designed to assess the relative effectiveness of three methods of coordinating programmed instruction with teacher instruction. A program, unlike workbooks or other tools available for the teacher's use, is constructed so that it will teach new material rather than simply supplement initial instruction given by the teacher. This capability provides the teacher with a versatile tool which he can use either to introduce new material for original learning or to provide relearning, review, and practice of new material which he himself first introduces. Although such versatility permits a number of possible teacher-program combinations, there are no data presently available to indicate the relative effectiveness of various arrangements of teacher and program instruction and review in an intact class situation. The purpose of this investigation was to compare three teacher-program combinations in terms of the amount of learning produced by each over a standard period of time. The subject matter taught in this study was single-digit addition and subtraction.

---

[1] In this chapter, individual studies will be designated by a number indicating the grade level and a letter indicating the specific study. Thus (1A) is a designation for Study A (Teacher-Program Coordination) performed at the Grade 1 level.

In the first combination, the teacher gave the original instruction in a small number of new facts that were to be learned, and followed this initial instruction with a program assignment which provided relearning, review, and practice with those facts. She then introduced more new material, and the cycle continued. In the second combination, a reversal of the first, the teacher assigned the program to do the initial teaching of the new material, followed this by additional teacher-directed review and practice of that material, then assigned the program again to introduce more new material. A third combination employed was assignment of the entire program to provide the sole instruction in addition and subtraction over a period of several weeks followed by an equal time period in which the teacher conducted daily sessions of relearning and review of the material originally introduced by the program.

Distribution of Practice

The purpose of the second investigation (1B) was to examine the effect upon learning from programmed materials of certain instructional variables which the teacher often manipulates in traditional classroom teaching. The subject matter in this study was time-telling.[2] One variable investigated was the spacing or distribution of learning sessions. Laboratory studies have indicated that, under certain conditions, learning is facilitated if practice is distributed over a period of time, with rest periods between practice sessions (McGeoch & Irion, 1952, Chapt. 5; Underwood, 1961). In the laboratory investigations, the periods between practice sessions have ordinarily been of rather short duration. In the present study, three different time periods of considerably longer duration than those ordinarily employed in the laboratory were used to separate learning sessions. One of the classes involved received two learning sessions a day with the time-telling program (morning and afternoon), making the daily interpolated intervals approximately one-half of a school day in length and permitting the class to complete in seven days a program which ordinarily requires 14 days of single sessions. A second group received a single learning session each day for 14 school days, making the interpolated interval between practices approximately 24 hours (with the exception of the two weekends that occurred in the course of the experiment). The third group was given a learning session every other day until the program was completed, making the interval between sessions approximately 48 hours (again with the exception of weekends), and the total time for program administration 28 days. A recent study (Reynolds and Glaser, 1963) employing spaced practice in a programmed instructional sequence has further suggested the possible facilitation of learning as a result of distribution of practice.

Pre-familiarization. Another teaching variable studied in the time-telling experiment was pre-familiarization with the material to be learned. There is evidence from verbal learning investigations that familiarity with the stimuli and responses that are to be associated facilitates learning of the associations (Noble, 1961, p. 221). Research with laboratory learning tasks other than verbal associates has also suggested that preliminary exposure to material that is to be learned facilitates later learning by arousing the learner's curiosity (Berlyne, 1960). These laboratory findings receive additional support from the well-established tendency of the classroom teacher to initiate new topics of study by presenting an "overview" of what is to be learned, presumably motivating students to find out more about the new topic. These several leads all indicated that if the time-telling program were preceded by activity sessions in which the teacher familiarized the children with various kinds of clocks, the importance of being able to tell time, etc., learning from the program might be facilitated. In two of the six first-grade classes which took the time-telling program, a week of pre-familiarization activity was given before the program was introduced.

---

[2] Glaser, R., Reynolds, J. H., and Weinstein, I. P. Time-Telling: A Teacher-Student Programmed Sequence. University of Pittsburgh, 1963. This program differs from the usual self-instructional sequence in that the teacher is an active member of the instructional procedure. The teacher and the class as a group work together at the beginning of each new unit and then students are permitted to work by themselves. The characteristics of this program are described in Glaser and Reynolds (1964).

Post-learning Practice. The final variable investigated in Study 1B was the effect of post-learning practice upon retention of what was learned in the program. All of the classes used in the study were given two tests--a test immediately following the program to assess learning, and a retention test two weeks later. Two classes were given daily five-minute practice sessions in telling time, under teacher direction, during the two-week interval between testings. One of these classes had also received prefamiliarization, but the other had not. The remaining classes received no formal post-learning practice in time-telling during the forgetting interval. It was expected that the use of short but consistent practice periods would not only facilitate retention, but provide additional learning as well for those learners whose performance on the first test showed this to be necessary.

Comparison with Non-programmed Classes

The third area of investigation (1C) in the first grade was a general comparison of the arithmetic achievement of the experimental programming classes during the school year with classes receiving arithmetic instruction by traditional methods only.[3] The experimental classes did not receive all of their arithmetic instruction throughout the year by programming methods. However, a substantial portion of the total time spent learning arithmetic did involve programming, or teacher efforts coordinated with programming, so it was of interest to determine the extent to which these instructional procedures affected overall class achievement for the year.

General Procedure

Prior to the beginning of the school year, six first-grade classes were selected for participation in the programmed learning studies. Since little standardized test data were available regarding individual differences before school began, an arbitrary choice was made of two classes from each of three elementary schools in the Baldwin-Whitehall school district. Neither the teachers nor any of the students in any of the classes selected had previous experience with programmed instruction. All of the teachers were experienced in teaching first grade, however, and were selected in part because of their previous success and positive attitudes toward teaching. Mean IQ scores and standard deviations on the California Test of Mental Maturity (Primary), and the size of each of the intact classes participating (after elimination of some students in the course of data analysis because of sustained absence or school transfer) are presented in Table 1.1. This table describes the groups used in the analyses made for the exploratory study and for Studies 1A and 1C. Slightly different groups, to be discussed later, were used in Study 1B.

Each class received the introduction to numbers program presented through the Min-Max II Teaching Machine at the beginning of the school year, during the six-week period from September 10 through October 19.

---

[3] By "traditional methods" is meant the usual teacher-classroom procedures being carried out by the particular school system at the time programmed instruction was introduced. The comparison made was between existing methods and a new technique. The term "existing methods" obviously is not rigorously defined. More definitive studies of the variables involved in the instructional process are needed in order to indicate the measurable dimensions on which two such methods vary.

Table 1.1

Group Size and Intelligence Means and Standard Deviations for the Six First-Grade Experimental Classes

Class	N	Intelligence	
		$\bar{X}$	s
U	21	114.00	15.05
V	23	110.57	13.42
W	17	104.29	10.13
X	17	102.88	12.74
Y	25	113.32	14.60
Z	18	111.06	14.44
Total Group	121	X=109.84	X=14.24

The explicit method of presentation and the results obtained are described in detail below in the Exploratory Study. Following the introduction to numbers program, the classroom teachers presented a review of numbers, certain measurement concepts (such as pints, quarts, one half of a whole, etc.), and readiness activities for addition and subtraction until December 10. On this date, all six classes began a ten-week period of coordinated teacher and programmed instruction in addition and subtraction facts. During this time, the classes were given two 20-minute arithmetic periods each day, one in the morning and one in the afternoon. Programmed or teacher instruction was presented in these sessions according to varying experimental conditions that are described in detail below for Study 1A.

Between March 18 and May 1, the teachers gave further practice and review in addition and subtraction, and instructed the classes in other non-programmed arithmetic topics required at the first grade level, e.g., a unit on money and counting by one's, five's, and ten's. In the last week of this period, two of the classes (randomly selected from the original six) were given additional activities which constituted pre-familiarization for the time-telling program. On May 1, all six classes began working in the time-telling program and proceeded according to the methods described under Study 1B below. Upon completion of this program, the teachers continued giving instruction in the usual Grade 1 curriculum topics for the remainder of the year. The schedule for the year is outlined in Table 1.2.

Comprehensive achievement testing, the results of which are described in the next section, was conducted before and after the administration of each program. Also, a comprehensive test of arithmetic knowledge was given at the end of the school year to all six experimental classes and to two traditional (control) classes. Comparisons between the experimental and control groups are presented in Study 1C below.

To insure that these complex and extensive procedures were carried out according to plan, prior to the beginning of the school year a comprehensive Teacher's Manual which served as a full-year's lesson plan was written by the elementary arithmetic supervisor in cooperation with the experimenters. This manual described the arithmetic procedures to be used on each school day, including such information as test dates, the activities and materials to be used on teacher instruction days, the specific kinds of readiness, pre-familiarization, and review procedures to be employed, etc. Throughout the year the experimental teachers followed these daily lesson plans, assisted and monitored by the elementary supervisor and the full-time research coordinator. Periodic meetings of the teachers, school supervisors, and experimenters were held to discuss progress, clarify ambiguities in the instructions, and reorganize or elaborate the manual as necessary.[4] By these procedures, adequate control was maintained over all experimental conditions throughout the year in terms of materials presented and manner and time of presentation. Of equal importance, the same degree of control was imposed upon the arithmetic activities interpolated between the specific experiments, providing some assurance that no group received distinct advantages or disadvantages from exposure to differing extra-experimental materials or methods. The obvious limitation in this control was the individual differences among teachers. Wherever possible, more than one teacher was involved in each of the experimental conditions in order to minimize the effects of these teacher differences.

Exploratory Study:
The Classroom Use of Simple Machines with Young Children

Method

In order to determine the utility of machines in presenting programs to groups of young children, six classes were administered the introduction to numbers program. Prior to beginning work on the program, teachers conducted four daily sessions in number readiness activities, which consisted of fingerplays, counting songs, or simple number games; none of these involved actual instruction in recognizing or writing numbers. Next, five periods of practice with the machines were initiated in which teachers showed students how to manipulate the machine knobs, draw circles around correct responses to simple figure-matching frames, and confirm the responses made. Following this pre-training period, a pretest was administered to each student individually to assess his existing

---

[4] A major reorganization of the experimental procedure did in fact occur as the result of one of these meetings. Originally, the plan was to administer the time-telling program in January, to be followed by the addition and subtraction facts program. The teachers indicated, however, that the entering behaviors necessary for the time-telling program, such as counting by five's and writing numbers from one through sixty, would not be established by that time. Consequently, the manual was rewritten to specify daily operations for the plan outlined in Table 1.2.

Table 1.2

Schedule for First Grade Arithmetic Instruction

September 10-September 14	Teacher instruction: number readiness
September 15-September 21	Practice with teaching machines
September 24-October 19	Introduction to numbers program
October 22-November 23	Teacher instruction: number review, measurement
November 26-December 8	Teacher instruction: addition and subtraction readiness
December 10-March 15	Addition and subtraction program and teacher instruction
March 18-April 18	Teacher instruction: review of addition and subtraction, money, measurement
April 21-April 29	Teacher instruction: counting by five's and/or pre-familiarization for time-telling program
May 1-May 20	Time-telling: a teacher-student programmed sequence
May 21-June 3	Teacher instruction: arithmetic review or time-telling post-learning practice

knowledge of numbers. The test was 40 items in length; Part I (10 items) measured the student's ability to write the numbers 1 through 10, Part II (10 items) measured the student's ability to count objects and then write the number, Part III (10 items) required the student to respond orally to printed numbers, and Part IV (10 items) required the student to copy printed numbers with a pencil.

On the day following the pretest, students began working with the program in daily 20-minute sessions. At the beginning of a session the student would take his machine from the shelf or table at the back of the room to his seat and wait for instructions. (By pre-loading the machines and taping names on them, the teachers ensured that machines could be returned to storage at the end of each session and be ready for continued use by the same student during the next session. Thus problems in loading and finding places every session were avoided.) On each of the first three days, the teacher guided students frame by frame for just 20 frames, explaining for each frame what was required and the steps the student must take to work successfully. From the fourth day on, students were permitted to work at their own paces, with the teacher circulating about the room to give encouragement and to help when machine problems developed. In this phase, however, no instructional aid concerning the context of the program was given. As each student finished, he was individually administered the numbers test again. Students who finished before the total time allotted for the experiment were assigned to practice and enrichment activities in number work after they completed the posttest. In this manner, all students completed the program over the four-week period allotted, working 20 minutes each day, and most had opportunity to engage in varying amounts of individual enrichment activities as well. The distribution of the number days (sessions) required by the children to complete the 615-frame program is shown in Figure 1.1. Figure 1.2 shows a scatter diagram of posttest scores of time to complete the program and final score. There appears to be little correlation between the two measures.

Results

Achievement. Means, SD's, and mean percent correct for the pretest and posttest, as well as gain scores for all groups, are presented in Table 1.3. The mean percentages on the pretest indicate that five of the six groups knew practically three-fourths or more of the material to be learned before being exposed to instruction. Even the groups with the lowest pretest percentages, which were also the lowest IQ groups (see Table 1.1),

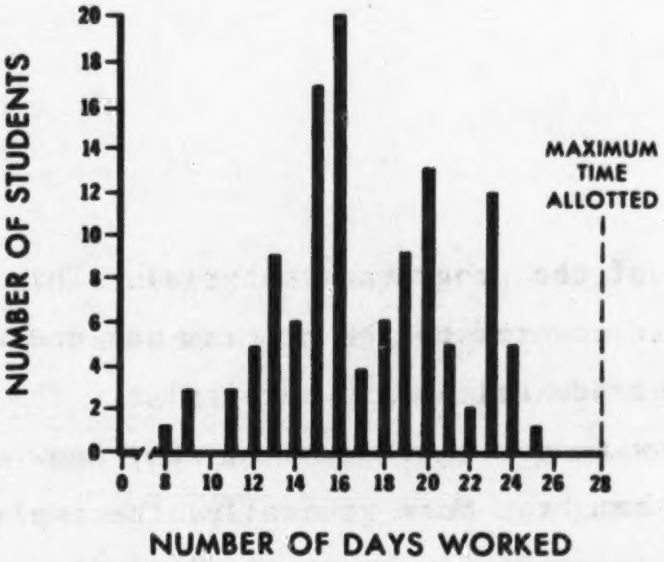

Figure 1.1 Distribution of number of days to complete introduction to numbers program (N=121).

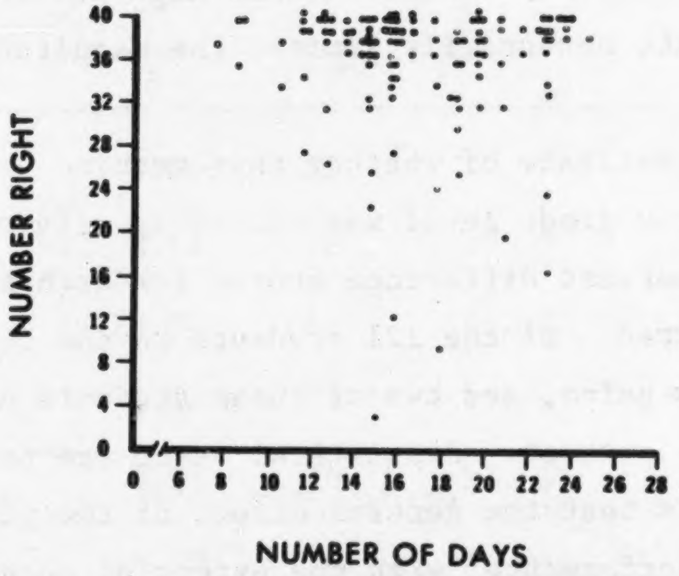

Figure 1.2 Scatter plot of final test scores and number of days to complete introduction to numbers program (N=121).

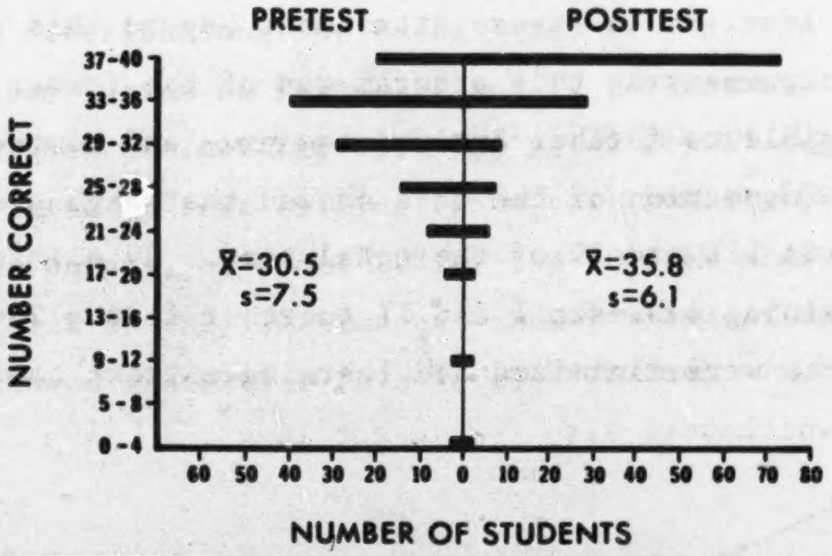

Figure 1.3 Pretest and posttest distributions on introduction to numbers test (N=121).

already knew over 50% of the programmed material. This finding is of interest, since the topic covered by the program was one which is usually included in the first-grade arithmetic curriculum. The implication is that in an intact class many children may already know a significant amount about areas which are taught. More generally, the implication is that pretesting of student achievement may be as significant a measurement procedure for effective instruction as is end-of-course examination. Frequency distributions, illustrated in Figure 1.3, show that 62% of the 121 students participating in the experiment reached the ceiling (or within three points of the ceiling) on the posttest. (This high achievement, plus the high pretest achievement, necessarily limited the magnitude of the gain scores.)

To obtain some estimate of whether this machine use of programmed instruction at this early grade level was generally effective in producing learning, pretest-to-posttest difference scores for each student in the six groups were calculated. Of the 121 students in the combined groups, only six failed to show gains, and two of these students had scores at the test ceiling on the pretest. Statistical tests are not necessary to conclude from these data that the general effect of the program was an improvement in number performance, with the extent of observable improvement being restricted in 62% of the cases by the low ceiling on the measuring instrument used. The existence of a test ceiling, however, poses some dilemmas. A "test ceiling" can be considered a "level of mastery" which it is desired that students attain and which is the instructional objective of a program. A test can be constructed either to assess the attainment of this level or to assess attainment beyond this point. Apparently the test accompanying this program was of the former type.

It was possible to further analyze the pre- and posttest means on certain subtests. Inspection of the data showed that high pretest scores predominated in Parts III and IV of the total test. By eliminating these two parts, the remaining subtests I and II together form a 20-item test in which ceiling effects were minimized and learning effects could be indicated.

Table 1.3

Means, SD's, Percentages Correct, and Mean Gains for Six First Grade Groups on the Pre- and Posttests for the Introduction to Numbers Program

Group	Pretest (40 items)			Posttest (40 items)			
	$\bar{X}$ No. Correct	s	$\bar{X}$% Correct	$\bar{X}$ No. Correct	s	$\bar{X}$% Correct	$\bar{X}$ Gain*
U	32.50	4.04	81	37.57	2.36	94	5.07
V	31.35	8.14	78	37.00	5.29	93	5.65
W	25.94	10.94	65	30.62	11.10	77	4.68
X	29.03	6.54	73	33.74	5.50	85	4.71
Y	31.44	7.21	78	37.26	3.05	94	5.82
Z	31.42	4.49	78	37.14	2.95	93	5.72

*number of items

Table 1.4 given the means, SD's, and mean gains of the six groups on Parts I and II combined and on Parts III and IV combined. Figure 1.4 shows the pre- and post-score distributions for Parts I and II and Figure 1.5 for Parts III and IV. The lower gains on Parts III and IV reflect the ceiling effect imposed by the higher pretest scores. The gains for I and II combined, which are less restricted by test ceiling, show that the six groups gained an average of four points, or about 20% of the 20-item measure, and that while on the pretest 12% of the 121 students showed at least 90% mastery, on the posttest 50% showed mastery at this level.

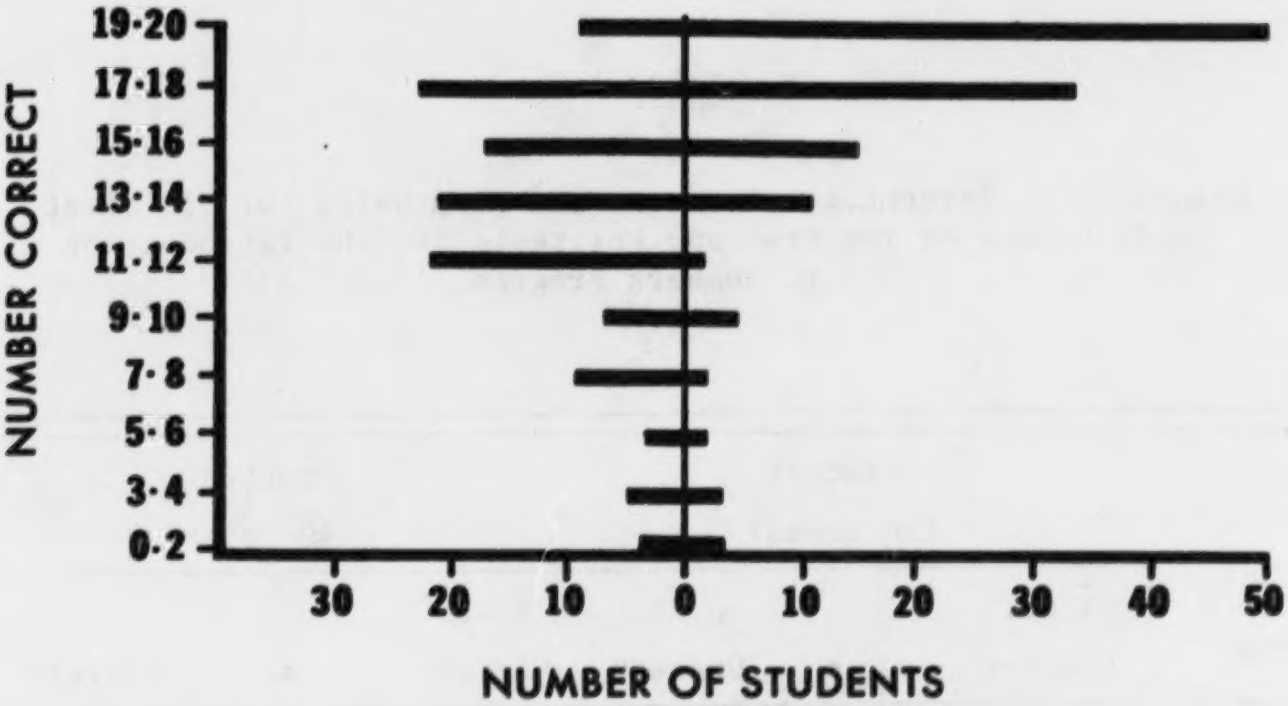

Figure 1.4 Pretest and posttest distributions on introduction to numbers program, parts I and II (N=121).

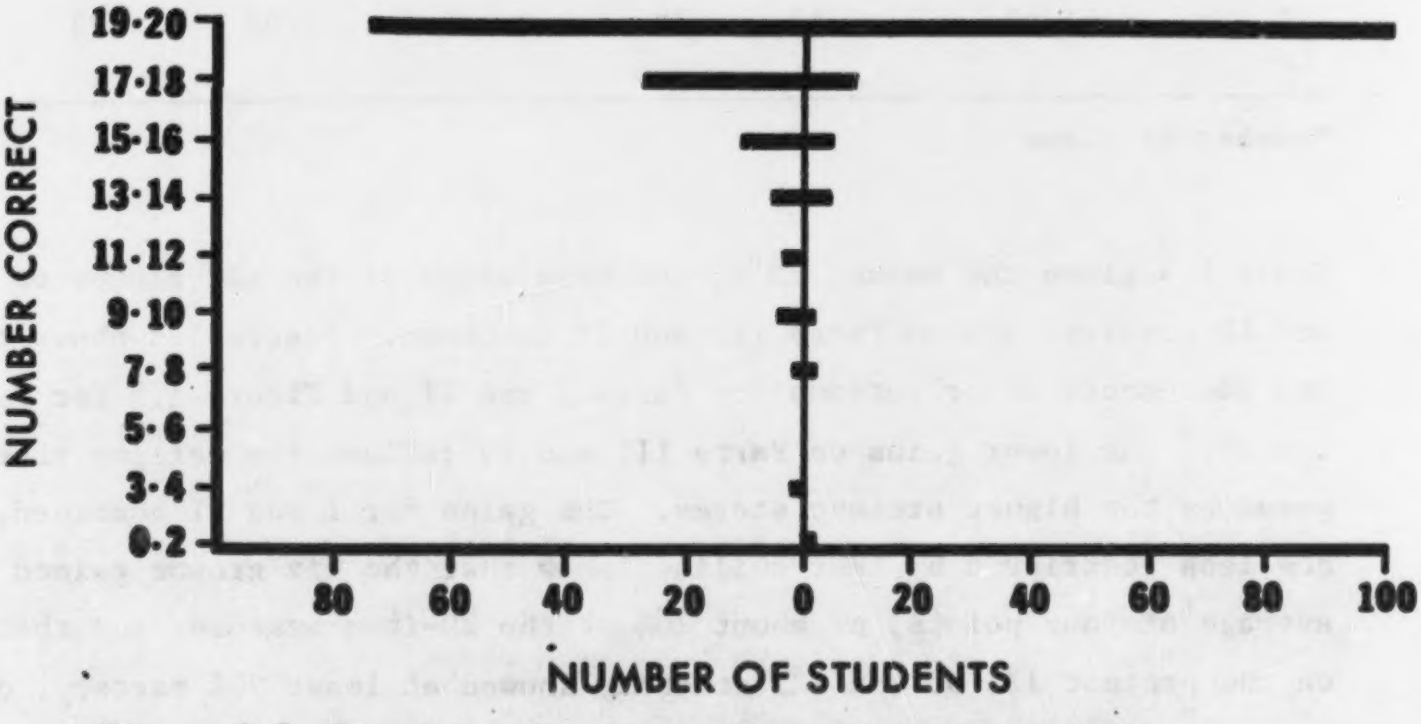

Figure 1.5 Pretest and posttest distributions on introduction to numbers program, parts III and IV (N=121).

Summary

Results of the exploratory study in which a simple teaching machine was used to introduce numbers to first-grade children indicated that these very young children could use and learn from instructional programs presented in machines. The particular machine used required the children to turn a knob and circle an object with a pencil. This was often a difficult task and it would seem that at the first-grade level the development of machines which require less motor skill on the part of the student while at the same time using his already obtained-skills is a highly desirable direction for the future.

This study with this particular program required both pre- and posttesting of the subject matter to be learned. Pretesting indicated the attainment of mastery on the part of a sizeable number of students, indicating that the use of the program for many of the students was redundant. While at the beginning of the program 59 out of 121 students indicated near mastery of the subject, after the program 102 out of 121 indicated such near mastery. However, the evidence provided by the pretesting indicates that individualized instruction based on the pretesting of attained competences prior to learning is a significant aspect of efficient instruction.

Effect of Intelligence. The relatively lower gain that is shown in Table 1.3 for Group W, which was the lowest of the six groups in mean intelligence, suggests that amount of gain on Parts I and II was related to intelligence. However, the correlation for the total group between intelligence and gain on I and II combined was $r=.13$, which is not significantly greater than a zero correlation for $N=121$. Further inspection of the relationship between intelligence and performance on Parts I and II combined showed that IQ, as measured for these young students, was a relatively poor predictor of either pretest performance ($r=.34$), or posttest score ($r=.28$). Intelligence, as measured, apparently had only a slight relationship to the task of learning numbers, and had no relationship at all to the gain accruing from programmed instruction in this task.

Table 1.4

Means, Standard Deviations, and Mean Gains of Six Grade 1 Classes on Parts I and II and Parts III and IV of the Introduction to Numbers Test

	PRE I and II (20 items)		POST I and II (20 items)			PRE III and IV (20 items)		POST III and IV (20 items)		
Groups	$\overline{X}$	s	$\overline{X}$	s	Mean Gain	$\overline{X}$	s	$\overline{X}$	s	Mean Gain
U	13.50	3.37	17.81	1.97	4.31	19.00	1.07	19.76	.61	.76
V	13.26	5.41	17.57	3.80	4.31	18.09	3.16	19.43	1.56	1.34
W	10.76	5.98	13.74	6.31	2.98	15.12	5.53	16.88	5.12	1.76
X	11.68	4.66	15.18	3.89	3.50	17.35	2.40	18.50	1.83	1.15
Y	13.36	3.90	17.54	1.92	4.18	19.28	1.37	19.72	.83	.44
Z	13.08	3.11	17.58	1.92	4.50	18.33	1.67	19.56	1.23	1.23
Total Group	12.72	4.60	16.73	3.89	3.92	18.01	3.12	19.08	2.46	1.07

Teacher Reactions. The general feeling of the participating teachers, as judged from informal conversations, was that the program had provided an adequate basis for further work in arithmetic. There was, however, some question about the relative efficiency of teaching the particular topic by program as compared with the traditional manner of teaching numbers, a question which was not investigated in this experiment. The most important and general criticism was that, although the students obviously learned, they were not adept at using the machines independently. Their undeveloped motor skills at this early age level resulted in inaccurate recording of responses, and frequent jamming of the machines. Initially it was necessary for the teachers, themselves inexperienced in machine usage in the classroom, to spend considerable time adjusting machines because the children were incapable of doing this independently. The problem diminished somewhat as the experiment progressed, but most of the teachers considered that the time consumed in this mechanical activity would have hindered their effectiveness had they been asked to give individual instructional help during the program sessions. The implication was that efficiency in programmed instruction at lower grade levels might be improved by presenting materials through machines which require less motor skill on the part of the student (e.g., automatic filmstrip displays, pressing buttons for responding, etc.) than those used in the present experiment.[5]

---

[5] It should be noted that the reported machine problems were probably not due to the machine itself. Older children, being less careless and more coordinated than first graders, have used the same machines over long periods of time without jamming and attendant problems noted here. Even the first graders had fewer problems once they became accustomed to the machine.

Study 1A:
The Effect of Varying Teacher-Program Combinations Upon Learning

Method

In order to employ and compare three different modes of teacher-program coordination, each two of the six original arithmetic classes were assigned to one of three conditions. The three condititions differed in the mode of instruction used to present new material and review material in addition and subtraction. Classes U and V, both located in the same school, constituted Group T-P; they received initial instruction in an addition and subtraction topic from the teacher (T), followed by practice and review in the same topic from the program (P), before advancing to a new topic. Classes W and X, in another school, formed Group P-T. This group received initial instruction in each new topic from the program, followed by practice and review under teacher direction before going on to new material. Finally, Group P-P, composed of classes Y and Z in a third school, received only the program during the first half of the experiment, working daily in the program until it was completed; after this the teachers in Group P-P reviewed all of the addition and subtraction facts daily for the remainder of the experiment.

Each condition received two 20-minute periods of arithmetic instruction per day, a period in the morning and a period in the afternoon, in order to facilitate the coordination between teacher and program for the T-P and P-T groups. For short topics, instruction could be given in the morning session, followed by review and practice in the afternoon. With topics requiring more than one session of initial instruction, consecutive program sessions were used, followed by a number of teacher sessions for practice and review for the P-T group, or consecutive teacher instruction sessions followed by a number of program sessions for the T-P group. (It was often necessary to allow more sessions for initial teacher instruction in the T-P group than were required for initial programmed instruction in the same topic for the P-T group.)

One week before the experiment began, teachers in the three groups conducted certain readiness activities considered to be prerequisite to starting the addition and subtraction facts program. The activities, taught only by the teacher, included review of number meaning and relative number value, counting practice, and learning to use new terms and signs (e.g., plus and minus, take away, etc.). Following this orientation week, a 90-item pretest was given containing 45 addition and 45 subtraction facts. For ten weeks thereafter, the three groups received coordinated teacher and program instruction in addition and subtraction, according to the experimental conditions just described.

Of the 25 units in the program, only the first 10 were teaching units, the last 15 being review and practice. The groups were paced through the program (i.e., told by the teacher exactly how many frames were to be completed during a given work period) on the teaching units. On the practice units 11-25, students were permitted to work at their own paces during the 20-minute sessions allotted to programming, but still worked as a group with the teacher in the teacher sessions. This individualized pace in the last 15 units of the program resulted, of course, in students finishing the program at varying times according to differences in individual speed and learning progress. As each student finished, he was given the 90-item posttest, which required an administration time of approximately 30 minutes, and then was given individual arithmetic assignments to complete daily until the other students had completed the program. The posttest was the same as the pretest and consisted of single-digit, two-row addition and subtraction problems.

Since the P-P Group received the entire program before being given any teacher-directed review, the testing procedure for this group varied slightly from that for the others. As students in the P-P group finished the program they were given a first posttest to assess what had been learned from the program alone. The second posttest was given following the subsequent teacher-directed review of addition and subtraction, which measured the achievement effect of the program-plus-teacher treatment.

The time and order of testing and instruction were specified for each group in the Teacher's Manual. Directions for administering all teacher-directed sessions, including topics to be covered and specific materials to be used in presenting either new or review topics, were also specified on a day-by-day basis. In this manner, the amount and type of material presented by the teacher was controlled over all groups.

Results

Achievement. One aspect of the experimental design did not proceed according to plan and should be taken into account in interpreting the achievement data. Three weeks after the beginning of the experiment the two teachers in the T-P group requested that they be permitted to give only teacher instruction to their classes for one week, rather than supplementing initial learning with the program review assignments as originally planned. This request was granted, and therefore data for this group is not wholly the result of the T-P procedure described above; rather, the data includes one week out of the ten in which teacher instruction was used without any additional program review. Reasons for this request and change will be described later.

Intelligence data and the various pre- and posttest scores for all groups are presented in Table 1.5. Figures 1.6, 1.7, and 1.8 show the pre- and posttest distributions for the three experimental groups. Since the P-T group was lower than the others in intelligence, and all three groups differed in pretest performance, statistical analyses of achievement were made for gain scores rather than posttest performance. On the total test, mean achievement gains for the T-P, P-T, and P-P groups were 44.51, 41.09, and 46.02, respectively. Comparisons between groups on these mean gains for the total test yielded $\underline{t}$ values ranging from .29 to .98 ($p > .10$), indicating that all group differences in gain were within chance limits. Comparisons of gains on the addition and subtraction subtests separately yielded similar results, with none of the differences among the groups large enough to reach statistical significance. Consistently, the varying experimental treatments failed to produce variations in amount of achievement gain. However, the fact that the T-P group achieved nearly the test ceiling (the mean percent correct for group T-P on the total posttest was 97%) indicates that many students in this condition actually mastered the subject matter while those in other groups did not. Whether mastery can be attributed to the experimental T-P treatment itself or simply to the higher pretest mean for the T-P group cannot be answered from the present data because the test was designed to measure the specific mastery to be attained, and the test had this level as a ceiling. Had the test included items requiring the student to use his knowledge for more difficult problems, it is possible that the T-P group would have demonstrated greater gains than P-T or P-P, which would indicate a superiority of the T-P treatment.

Mean scores and distributions on the interim posttest, taken by the P-P group immediately following programmed instruction and before beginning teacher instruction, are also shown in Table 1.5 and Figure 1.8. Examination of the addition and subtraction interim and posttest scores shows that the program alone was more effective in teaching addition than subtraction. For the total test, however, it can be seen that achievement gain made by group P-P following programming alone was approximately half as large as the gain following combined programming and teacher presentation in groups T-P and P-T. From these data it appears that the two instructional modes contributed about equally to the total achievement gain. To substantiate such an impression, however, it would be necessary to evaluate gains from teacher instruction alone during the first half of the total experimental period. Possible effects of the teacher alone, or the program following teacher introduction, cannot be determined from the present experiment. But it is clear from the interim posttest performance of the P-P group that exposure of first-graders to this particular program, without intervention by the teacher, did not produce the amount of achievement possible when teacher and program were coordinated over a longer period of time.

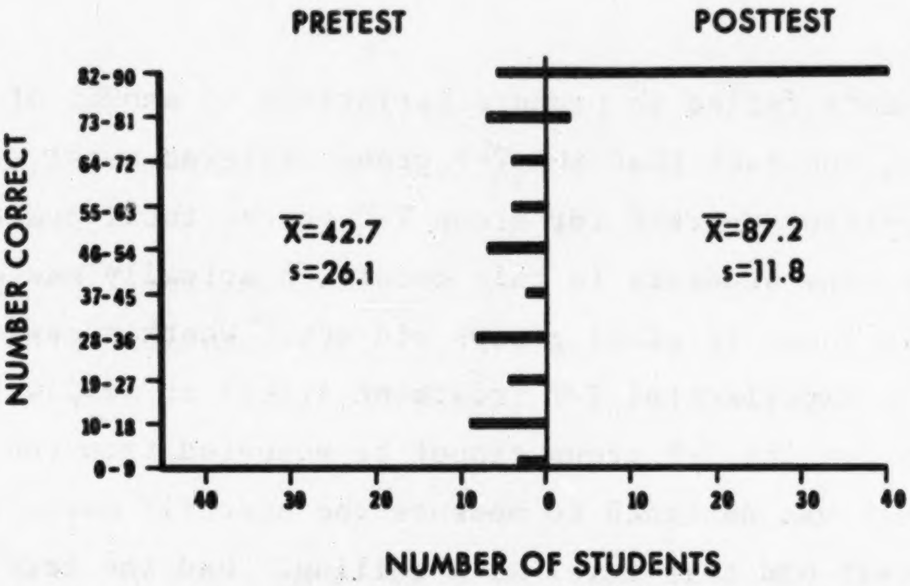

Figure 1.6 Pretest and posttest distributions for addition and subtraction, Group T-P (N=44).

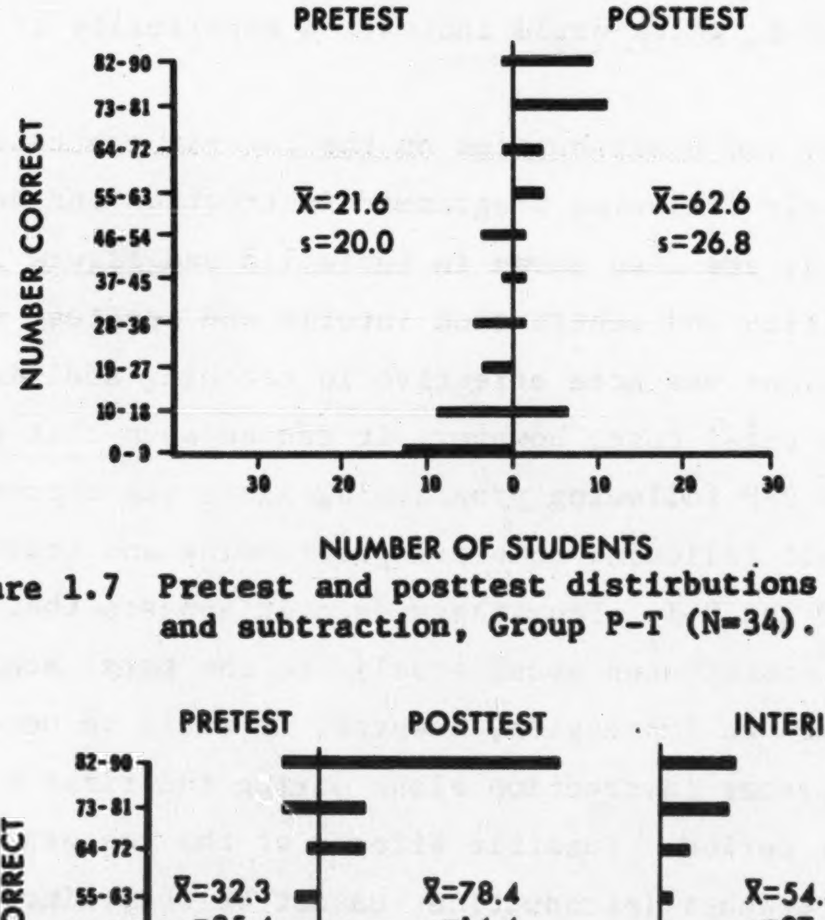

Figure 1.7 Pretest and posttest distirbutions for addition and subtraction, Group P-T (N=34).

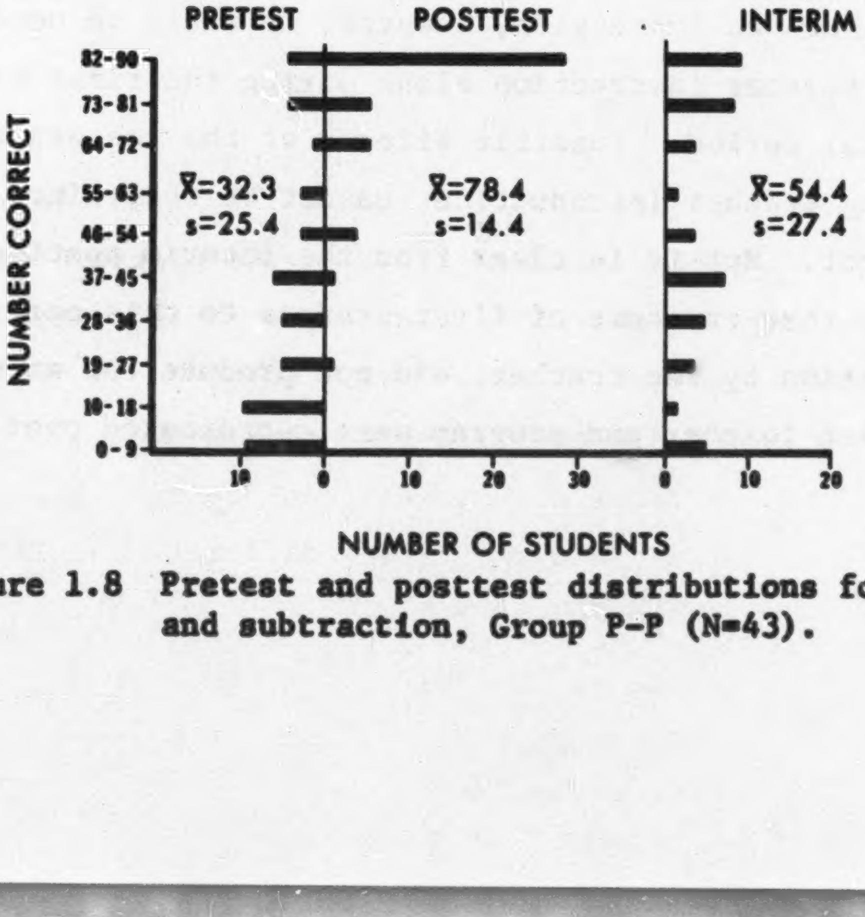

Figure 1.8 Pretest and posttest distributions for addition and subtraction, Group P-P (N=43).

Additional Comment. A general concern of the teachers was that the organization of this program was very different from that which was used when teaching this topic by traditional methods (for example, teachers felt that the program was not organized around "number families," the number line was presented at the wrong time, etc.). This "problem" precipitated a tempory deviation from the original design for the T-P group because the T-P teachers felt that they needed to spend a week re-teaching the material in a way that comformed more closely with methods they had used in the past. Following this brief period students continued with the original design, and no further problems were encountered that necessitated changing the design again. It is interesting that discrepancies between the program structure and the subject matter organization customarily used by the teachers in traditional instruction did cause some difficulty, since it demonstrates quite clearly the desirability of having programs evaluated for content and structure by the teachers planning to use them before they are put in use. It also emphasizes the fact that programming research and development may result in new ways of presenting the subject matter and that teachers may need to empirically determine for themselves, by classroom tryout or by the evaluation of field test data, the effectiveness of these new methods.

Summary

Interpretations about the use of a program must be made in the light of the characteristics of that particular program. The addition and subtraction program employed in this experiment, when used by itself for the P-P group, taught only to partial test mastery. This is indicated in Table 1.5 and Figure 1.8. With this particular program, teacher instruction increased the specified mastery level on the program test. In the context of this situation, the different arrangements of teacher-program coordination had no differential effect on student attainment. While the different groups achieved different levels, this appears to be primarily a function of the entering level of arithmetic achievement; average gains in achievement, however, were the same for the different experimental groups.

It is to be pointed out that the experimental variations made here were group (intact class) manipulations which worked with the class (all students) as the instructional unit rather than the individual student as the instructional unit. When the class is conceived as the instructional entity, instructional variations aimed at total class manipulations appear to have little effect. What seems to be necessary are variations in instruction aimed at the individualization of instructional procedures.

Study 1B:
Effects of Prefamiliarization, Distribution of Instruction, and Post-learning Practice Upon Programmed Instruction in Telling Time

Method

Design and Procedure. In the time interval between Studies 1A and 1B, the six first-grade teachers instructed the experimental classes in (a) counting and writing by one's from 0 to 60, and (b) counting and writing by five's from 0 to 60. Formal instruction in these skills ensured that nearly all students participating in Study 1B would learn the prerequisite behaviors necessary for working through through the time-telling program. Following this, each of the classes received the time-telling program under one of the six experimental treatments which differed in types of prefamiliarization, distribution of practice, or post-learning practice. The design is summarized in Table 1.6, which shows for each class the treatment received and the symbols which will be used to designate them. The design permits comparison of groups receiving the three levels of distributed practice (DP), i.e., programmed learning periods once a day (DP1), twice a day (DP2), or every other day (DPO); and also groups receiving various combinations of either prefamiliarization (PreF) or post-learning practice (PostL) with DP held constant at one programmed learning session per day.

The total program involved both paced group participation and individual work and required 14 instruction periods. Consequently, Group DP2 completed the program in just seven school days while Group DPO required

Table 1.5

Means, Standard Deviations, and Mean Percents for Intelligence Data, Pre- and Posttest Scores of the Three Groups in Experiment 1B

Group	N		I.Q.	Total Test (90 items)			Addition (45 items)			Subtraction (45 items)		
				Pre	Interim*	Post	Pre	Interim*	Post	Pre	Interim*	Post
T-P	44	$\overline{X}$	112.21	42.68		87.25	27.16		44.23	15.52		43.00
		s	14.33	26.06		11.82	13.33		.77	14.50		2.15
		%	------	47		97	60		90	34		96
P-T	34	$\overline{X}$	103.59	21.56		62.65	15.65		33.79	5.79		28.85
		s	11.53	20.00		26.85	13.43		12.67	9.20		14.89
		%	------	24		70	35		75	13		64
P-P	43	$\overline{X}$	112.37	32.33	54.40	73.35	23.26	33.53	41.14	10.93	20.86	37.21
		s	14.58	25.36	27.39	14.36	15.05	13.30	5.85	17.04	16.52	9.33
		%	------	36	60	87	52	75	91	24	46	83

*Administered to P-P Group only.

28 days to finish by working every other day. For one week prior to beginning the program, the PreF and PreF-PostL groups received a period each day in which the teacher presented various activities designed to acquaint them with clocks and arouse interest in telling time, e.g., making a scrapbook of clock pictures, drawing a clock face, and reading a story about clocks. The remaining groups proceeded with their usual class activities during this week. Following the program, the PostL and PreF-PostL groups received practice in reading a clock for ten minutes each day for two weeks, while other groups continued their usual work without further time-telling practice. The practice periods, which occurred during the time between a posttest and a retention test on the material taught in the program, consisted of the teacher showing the class various time settings on a large clock held up in the front of the room and asking the student to either write or say aloud the times indicated. After each setting, the teacher told the students the correct answer and wrote it on the blackboard. No additional instruction in how to read a clock was given during the post-learning periods, however.

Testing. All groups received a pretest the day before beginning the program, a posttest on the day following completion of the program, and a retention test two weeks after the posttest was administered. (Groups PostL and PreF-PostL received post-learning practice during this two-week interval.) The pretest, group administered by the teacher, required the student to write the numbers from 0 to 60 by five's, to write certain numbers between 0 and 60 as they were dictated by the teacher, and to write the times indicated by 12 small clocks (1 1/2" diameter) mimeographed on the test sheet and eight settings of a large Western-Union clock (11 1/2" diameter) displayed in front of the classroom. These tasks provided measures of each student's skill in the prerequisite behaviors necessary for taking the program, and also indicated the degree to which students already knew the material to be learned.

At both the posttesting and the retention testing, which were identical in content and procedure, students were presented with a series of clock settings, to which they responded by either writing or saying aloud

Table 1.6

Summary of Design for Study 1B

Original Group	Group Symbol for Study 1B	TREATMENT Prefamiliarization	Program Distribution	Number of Days to Complete Program	Post-learning Practice
U	DP1	No	Once a day	14	No
V	DP2	No	Twice a day	7	No
W	DPO	No	Every other day	28	No
X	PreF	Yes	Once a day	14	No
Y	PostL	No	Once a day	14	Yes
Z	PreF-PostL	Yes	Once a day	14	Yes

the time indicated by each setting. The first series consisted of eight printed clocks identical to the six-inch clock faces used in the program. This test, called the Program Clocks Test, contained the following settings: 2:47, 6:30, 11:24, 10:10, 3:21, 9:19, 12:30, 8:40. The second series were twelve smaller clock faces, the same as those used in the pretest, which were printed on a single test sheet. The settings (only to the five-minute mark) appearing in this Small Clocks Test were: 8:00, 11:05, 8:15, 7:45, 2:30, 11:50, 4:20, 6:15, 3:35, 10:55, 12:10, 5:40. For the third test, eight settings of the real Western-Union clock were shown to students by the experimenter who set the clock to each setting in turn and then held it up so the student could observe it and respond. The settings of the Real Clocks Test were: 1:05, 6:40, 4:00, 10:19, 3:52, 9:30, 12:47, and 8:04. These tests provided two measures of what was learned from the clock stimuli used in the program: a written measure, consisting of the total number of correct written responses the student made to all 28 clocks in the three tests combined, and an oral measure, which was the total number of correct oral responses the student made to the 28 clocks.

Both the posttest and the retention test required three administrators and approximately three hours time per group. Written tests were given first, on a group-testing basis, followed by individual testing of each student for oral responding. In the written test, the group was given the Program Clocks Test (in a stapled booklet) and instructed to "write what each clock says in the box underneath the clock." Then the Small Clocks Test was administered, with the same instructions. Finally, the tester displayed the various settings in the Real Clocks Test from the front of the room, and students wrote their answers in specially constructed booklets. Following this written phase, each student in turn was called from the room and asked to respond orally to the same three tests. In this phase, the first experimenter presented the Program Clock settings to the individual student and asked him to "tell me what these clocks say." The second experimenter presented the Small Clock settings in the same manner, and a third experimenter administered the Real Clocks Test. Each experimenter recorded verbatim all oral responses which occurred.

Results

Performances on the clock-reading sections of the pretest indicated that none of the 121 students could tell time prior to administration of the program. Seven students were eliminated because their pretest scores on the counting and number-writing sections were below 70%, indicating that they lacked the prerequisite behaviors necessary for taking the program. Thirty-two other students from the various groups were absent from school at critical times during the experiment, missing one or more of the tests administered, and had to be eliminated from the final analyses. These eliminations, the result of a run of colds and measles, caused drastic reductions in the size of the groups available for final comparison.

Pretests. Table 1.7 presents the N's of the groups eligible for final analysis, and the means and SD's of each group on the four sections of the pretest. Since the total possible scores on the numbers and counting sections were 12 and 10, respectively, it is apparent that students in all groups had the prerequisites for taking the program. The consistently low means on the clock reading sections indicates that no group was able to read a clock and write down the time prior to taking the program. Table 1.7 also presents the IQ means and SD's for the groups, which vary from those reported in previous experiments because of the elimination of many students.

Posttests. The six subtests administered immediately following the program were combined to yield two scores--a written posttest score consisting of the total correct written responses on the Program, Small, and Real Clock Tests, and an oral posttest score, consisting of the total number of correct oral responses on these three tests. The data for each group on the written and oral posttests are presented in Table 1.8. These data were used in testing two experimental hypotheses: (1) that prefamiliarization facilitates learning, and (2) that the amount learned from a programmed sequence is related to the manner in which practice sessions are distributed over time.

Table 1.7

Group Sizes and Pre-Test Means and SD's for the Groups Used in the Analyses of Experiment 1B

Group	N	IQ		Pretests							
				Writing Numbers		Counting by 5's		Small Clocks		Real Clocks	
				Total Possible-12		Total Possible-10		Total Possible-12		Total Possible-8	
		$\overline{X}$	s	$\overline{X}$	s	$\overline{X}$	s	$\overline{X}$	s	$\overline{X}$	s
DP1	18	114.00	15.56	11.89	.46	10.00	.00	.13	1.41	.08	.69
DP2	17	106.88	13.69	11.77	.73	9.82	.51	1.18	.71	1.18	.62
DP0	8	102.00	10.44	11.50	.87	9.75	.43	.38	.48	.38	.48
PreF	10	101.50	10.68	11.40	1.28	10.00	.00	.60	.66	1.00	1.00
PostL	20	111.40	13.71	11.75	.89	9.80	.68	.80	.40	.70	.56
PreF-PostL	9	106.80	14.90	12.00	.00	10.00	.00	.67	.67	.22	.42

Table 1.8

Means and SD's for All Time-Telling Groups on Written and Oral Posttests and Retention Tests

		Posttest				Retention Test			
		Written 28 items		Oral 28 items		Written 28 items		Oral 28 items	
	N	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
DP1	18	19.00	8.58	18.22	7.75	19.17	8.11	19.17	7.91
DP2	17	18.76	7.62	11.47	7.43	17.29	7.36	17.71	9.25
DPO	8	11.13	8.97	10.51	9.50	15.01	10.62	13.62	9.62
PreF	10	10.10	6.06	6.80	7.30	10.70	7.47	8.10	7.73
PostL	20	19.05	7.45	12.60	6.56	18.85	7.90	17.40	7.62
PreF-PostL	9	15.22	7.39	10.78	6.87	17.89	4.77	15.67	6.18

1. Prefamiliarization effects. The appropriate analyses for determining the effect of prefamiliarization are comparisons of Group PreF and Group DP1, each of which had the same learning treatment (program once daily) but varying familiarization treatments. Inspection of the means for these groups on the written and oral posttests indicates that, contrary to the hypothesis, Group PreF performance was consistently lower than that of Group DP1. The difference in means between these groups is significant on each posttest measure ($t=2.80$, df/26, $p<.01$, and $t=3.66$, df/26, $p<.01$, respectively).

However, analysis of mean IQ scores for the two groups indicates that the mean IQ for Group DP1 was significantly higher than that of Group PreF ($t=2.50$, df/26, $p=.01$). This would seem to indicate that the lower posttest scores for the group that received prefamiliarization might be not entirely due to differences in treatment but possibly to differing IQ levels, although in subsequent studies in this report IQ was found to be a poor predictor of post-program achievement scores.

A second analysis testing the effect of prefamiliarization is the comparison of Group DP1 with Group PreF-PostL, since the latter group had received the prefamiliarization treatment and the program once a day but at the time of the immediate posttests had not yet had any postlearning practice. An analysis of mean IQ scores indicated no significant difference between the two groups ($t=1.17$, df/25, $p>.05$). Although these two groups did not differ reliably on the written posttest ($t=1.06$, df/25, $p>.05$), the direction of the difference was again contrary to prediction. On the oral posttest, the difference contrary to the predicted direction was found to be significant ($t=2.35$, df/25, $p<.05$). These results are generally consistent with the DP1 vs. PreF comparisons above, but because of the difference between the groups in the first comparison it is difficult to state explicitly that prefamiliarization had an adverse effect upon posttest performance.

2. Distribution of practice effects. An analysis of variance performed on the IQ data for the DP1, DP2, and DP0 groups indicated that group differences in intelligence were not significantly greater than chance ($F=2.12$, df/2, 40, $p>.05$). Further variance analyses of these groups were performed for the written and oral posttest data to determine if the varying distribution of practice treatments produced differences in time-telling performance at the end of the program. Although the mean differences among the groups appear to be quite large in Table 1.10 the SD's are large also; consequently, the F's obtained for the two posttests were 2.69 and 3.73, respectively, neither of which is significant at the .05 level with df=2, 40. Apparently the distribution treatments of two sessions per day, one session per day, and a session every other day had no reliable effect upon posttest performance.

Retention Tests. Table 1.10 also summarizes the written and oral data obtained on the retention tests, which were given two weeks following the administration of the posttests. The hypotheses tested with these results were: (1) that prefamiliarization facilitates retention, (2) that post-learning practice facilitates retention, and (3) that retention is affected by differences in the distribution of practice sessions during learning.

1. Prefamiliarization effects. Inspection of Table 1.10 shows that Group PreF again performed at a consistently lower level than Group DP1, contradicting the original hypothesis. The differences between the two groups are significant on both the written retention measure ($t=2.62$, df/26, $p<.02$) and the oral measure ($t=3.45$, df/26, $p<.001$). A second comparison between Group PostL and Group PreF-PostL, was made to determine if the prefamiliarization treatment had any effect when combined with post-learning practice. An analysis of mean IQ scores showed the differences between the groups to be not significant ($t=.79$, df/27, $p>.05$). As can be seen in Table 1.10 the mean differences are again in the opposite direction from the one predicted. The group receiving prefamiliarization had lower means on both the written and oral retention tests, although the difference between these groups was not large enough to be considered reliable ($t=.33$ and .58, respectively, df/27, $p>.05$). However, an analysis of gain scores over the retention period indicates no significance in gain on oral responding ($t=.35$, df/27, $p>.05$), but a highly significant gain in written responding ($t=7.89$, df/27, $p<.005$), indicating that prefamiliarization may have some facilitating effect upon retention.

2. Post-learning practice. The evaluation of effects of post-learning practice was made by comparing Group DP1 with Group PostL, since both had received the same learning conditions (program once per day) but only the latter had received teacher-directed practice in telling time during the two weeks which elapsed between the posttests and the retention tests. The *t* tests performed showed no significant differences between those groups on either the written ($t=.15$, df/36, $p>.05$) or the oral ($t=.68$, df/36, $p>.05$) retention measures. However, an analysis of gain

scores shows that Group PostL demonstrated a significantly higher gain in achievement over the retention period ($t=2.22$, df/36, $p < .05$) on the oral retention measure, although the gain during the retention period on written responding was non-significant ($t=-.21$, df/36, $p > .05$). It is probable that the gain in oral responding was to some extent related to extra-class practice in time-telling. This would seem to indicate that gain over the retention period was little facilitated by post-learning practice.

It should be noted again that the post-learning practice did not include further instruction in time-telling, but simply a regular daily exposure to the taks already learned in the program. Nevertheless, the finding that such exposure had little effect upon subsequent performance was unexpected. The implication is that, with young children, practice with materials that are not already completely learned may not necessarily constitute opportunity for further learning through discovery or example. Instead, it appears that specific and well-analyzed instruction is the best procedure to insure learning to mastery. On the other hand, the data suggest that what has been learned at a concentrated rate may be retained at full strength by young children for a period of at least two weeks without any formal practice. Further research concerning the retention of first graders would seem to be most interesting and fruitful.

3. Distribution of practice effects. The F values obtained from analyses of variance performed with groups DP1, DP2, and DP0 on the written and oral retention measures were .66 and 1.03, respectively, indicating that differences among the groups were well within .05 limits of chance on both tests. The various distribution of practice schedules used during programmed instruction had no significant effects upon retention of either written or oral response learning. These results coincide with the non-significant findings obtained for the distribution of practice variable on the posttests. Apparently variations in instruction schedules, with this type of material at least, can be made quite freely in a classroom situation without affecting student achievement.

General Comment

Although the program did not teach all students to mastery, many were capable of reading a clock to the minute interval by the end of instruction. Skill at this level of precision was impressive to the teachers, because most attempts at teaching time-telling to first graders go no further than instruction for reading five-minute intervals, and relatively few children, at this grade level, learn to read a clock with even that much accuracy. Figures 1.9 and 1.10 show bar charts of the number of students reaching mastery or non-mastery in the two higher intelligence classes (DP1 and PostL) and the two lower intelligence classes (DP0 and PreF), respectively, on both the posttests and retention tests. The graphs indicate that the students in the two higher IQ classes generally achieved higher scores on the various final tests. The lower intelligence classes had a consistently smaller percentage of success, suggesting that IQ is probably related in some degree to mastery of this particular program. Although the performance is by no means perfect, the teachers felt that the program had taught many students and had provided considerable learning for many others, probably making further instruction to complete mastery far less difficult and time consuming than if the program had not been administered.

Summary

The time-telling program was built as an experimental programmed sequence to study certain aspects of learning in young children. It was used in this study as a means of teaching, rather intensively, a skill which is usually taught incidentally in the course of the first and second grades in connection with arithmetic topics. In contrast to addition and subtraction, time-telling represents a topic which teachers generally find difficult to teach and children find difficult to acquire, and therefore the teachers readily welcomed programmed instruction. The addition and subtraction program was less well received because teachers generally are quite successful in teaching addition and subtraction facts. The time-telling program taught many children to mastery but questions remain about

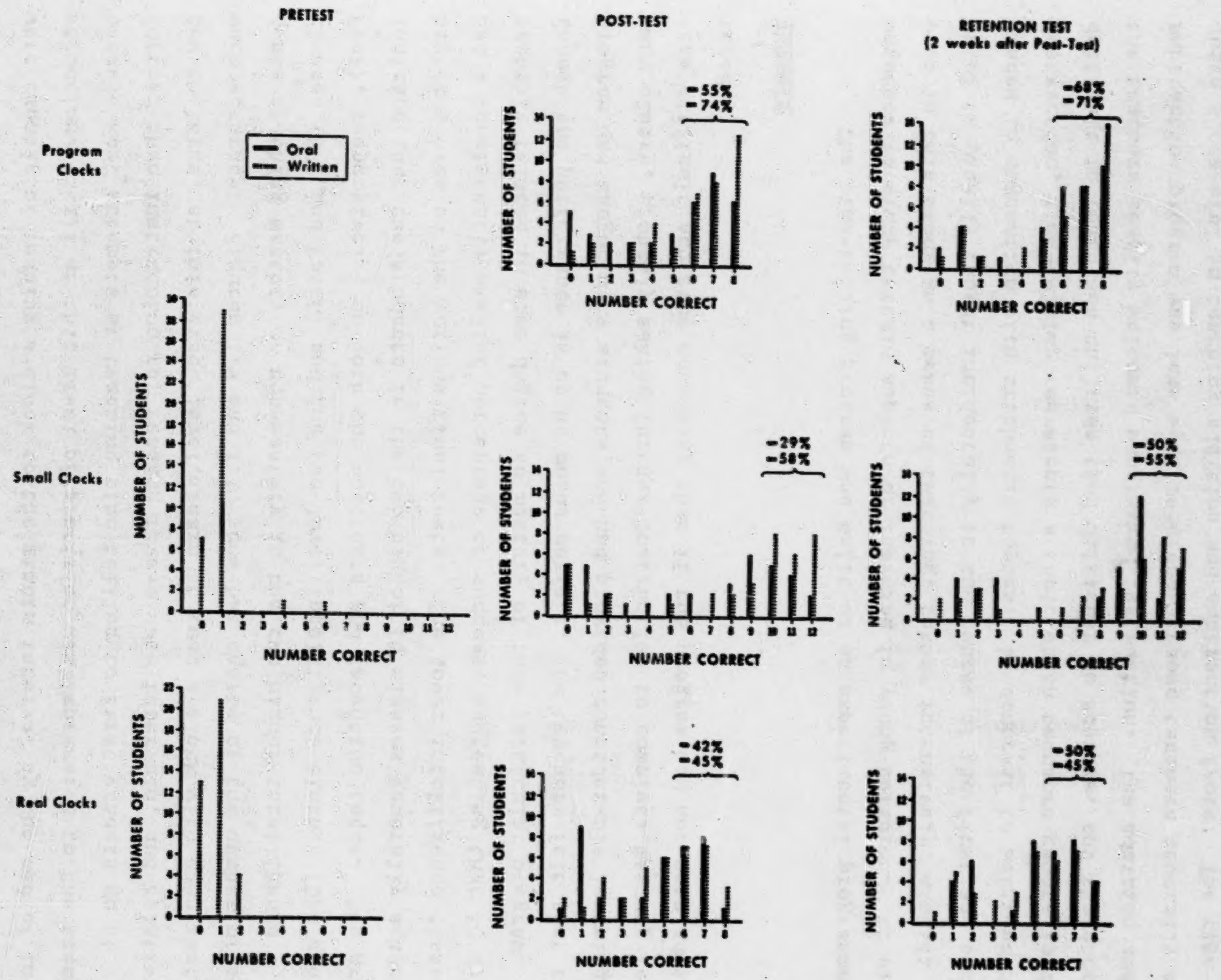

Figure 1.9 Pretest, posttest, and retention performance of students (N=39) in the two higher I.Q. first grades (average intelligence 112.36). The percent of students attaining 75% or more of complete mastery is recorded above each graph.

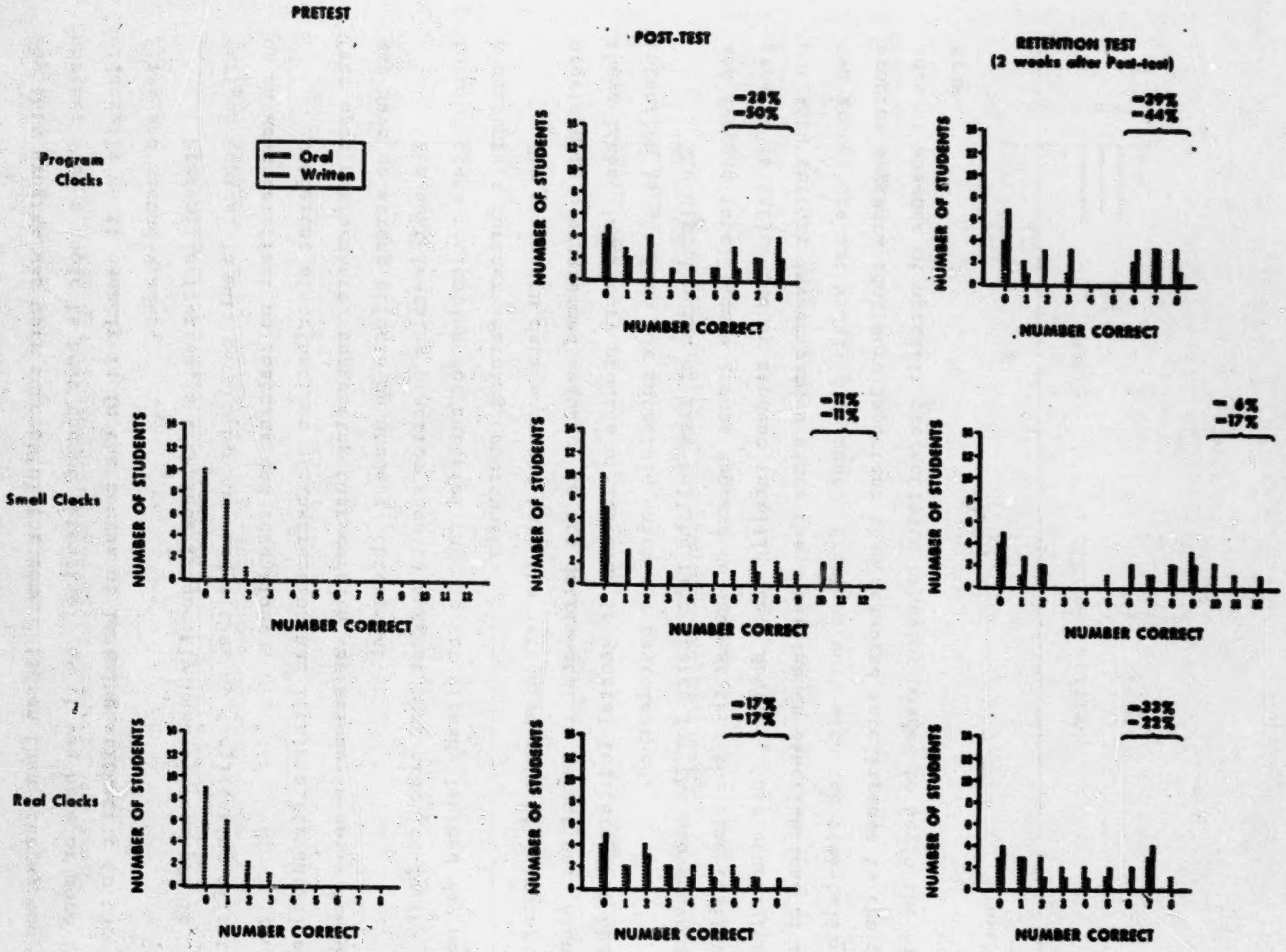

Figure 1.10 Pretest, posttest, and retention performance of students (N=18) in the two lower I.Q. first grades (average intelligence 101.75). The percent of students attaining 75% or more of complete mastery is recorded above each graph.

why this program was more successful for some children than others, and whether such a topic is best taught intensively as it was here or more incidentally as it usually is in the course of learning arithmetic in the first and second grades.

Prefamiliarization, a technique frequently used by teachers in various topics, in this particular case had either no facilitating effect or an adverse effect on learning and retention.

Different distributions of instruction had little differential effect upon performance, suggesting that certain variations in daily schedules may have no strong effects on student achievement.

The post-learning practice results suggest that practice in the form of class recitation for material that is not already learned may not constitute a further learning opportunity.

The retention data suggests that the intensive rate of learning provided in a programmed sequence is not detrimental to retention. Along these lines, laboratory studies of learning in general indicate that best retention is obtained for materials which are best learned.

The distribution of time-telling test scores for the above-average and average intelligence groups appears to suggest that the time-telling program is influenced by student intelligence. However, this conclusion warrants further investigation since the intelligence distributions of the two groups are not widely divergent. Previous work with the time-telling program suggests that more important than measured intelligence is the presence or absence of specific prerequisite behavior needed to begin the program.

Table 1.9

Means and Standard Deviations for Experimental and Control Groups on the Addition and Subtraction Test

Group	N	Total A and S Test (90 items)		Addition (45 items)		Subtraction (45 items)	
		$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
Program	121	82.02	13.99	42.08	5.03	40.33	8.71
Control	38	82.71	11.42	42.37	4.57	40.34	7.22

Table 1.10

Means and Standard Deviations for Experimental and Control Groups on the Written and Oral Time-Telling Tests

Group	N	Written (28 items)		Oral (28 items)	
		$\bar{X}$	s	$\bar{X}$	s
Experimental	121	16.13	8.47	14.72	8.59
Control	38	5.50	5.16	6.16	5.19

Written $t=7.28$, df/157, $p < .001$

Oral $t=5.78$, df/157, $p < .001$

Study 1C:
Comparisons of Program and Non-Program Groups at the End of Grade 1

Method

Two first grade classes which had received no programmed instruction during the school year were selected for comparison with the combined six experimental classes on a series of arithmetic achievement tests given at the end of the year. The non-program (control) classes were chosen so that their combined mean IQ was as equivalent as possible to the mean of 109.84 (s=14.24) of the total Program Group (N=121). The result of this selection procedure was a Control Group (N=38) with a mean intelligence of 108.97 (s=12.15) which had received traditional instruction in numbers and addition and subtraction from one of two experienced first-grade teachers, and had also received incidental teacher instruction in time-telling.

One week before school ended the Control Group was administered the following tests on successive school days: the Metropolitan Achievement Tests (Primary I Battery), which required five daily administration periods and included subtests of Word Knowledge, Word Discrimination, Reading, and Arithmetic; and the 90-item addition and subtraction test, described in Study 1A. The six experimental classes were administered the Metropolitan Achievement Tests and the addition and subtraction test during the same time period. Since these latter classes had taken the time-telling tests less than one week previously as retention measures for Study 1B, these tests were not readministered to them.

Results

Addition and Subtraction Tests. Since no differences were found among the T-P, P-T, and P-P groups in addition and subtraction performance in Study 1A, the three groups were combined into a single Program Group for comparison with the non-program Control Group. Table 1.9 presents the means and SD's of the two groups on the 90-item addition and subtraction test, and also on the addition subtest (45 items) and subtraction subtest (45 items) alone. Figure 1.11 shows the score distributions for each group on the total test. Statistical tests are unnecessary to conclude that there are essentially no differences between group means on any of the three measures. The Program Group, which had essentially the usual amount of teacher instruction plus the program, did only as well as the group receiving traditional instruction for the usual amount of time. Although many students in both groups achieved at a level close to mastery of the subject matter as measured by the test, scores greater than 90% being obtained by 80% of the Program Group and 73% of the Control Group, the extra time spent on the use of the program did not enhance learning of the addition and subtraction facts.

Time-Telling Tests. The means and SD's of the total Program Group and the non-program Control Group for the written and oral time-telling measures are shown in Table 1.10. The Program Group data includes students who failed the time-telling pretest but received the program, and also students who were not included in the preceding study (1B) because they were absent for one or more of the tests administered. Including these students lowered the test means to those reported in Table 1.10. Even with their inclusion, it is quite apparent that the time-telling performance of the Program Group was superior to that of the Control Group. Statistical analyses indicate that the mean differences were highly significant, with t values of 7.28 for the written measure (df/157, $p < .001$) and 5.78 for the oral measure (df/157, $p < .001$). The score distributions are shown in Figures 1.12 and 1.13. The data, however, indicate need of program revision since only 15% of the students obtained mastery (90% or better) on the end-of-year tests. The program was given on a group basis and revision of this procedure seems to be required.

Metropolitan Achievement Tests. Raw score and grade equivalent means and SD's obtained for the Control Group and the six experimental groups combined on each of the four Metropolitan subtests are presented in Table 1.11. Differences between the raw score means, analyzed by critical ratio tests, were within the 0.5 limits of chance for all of the

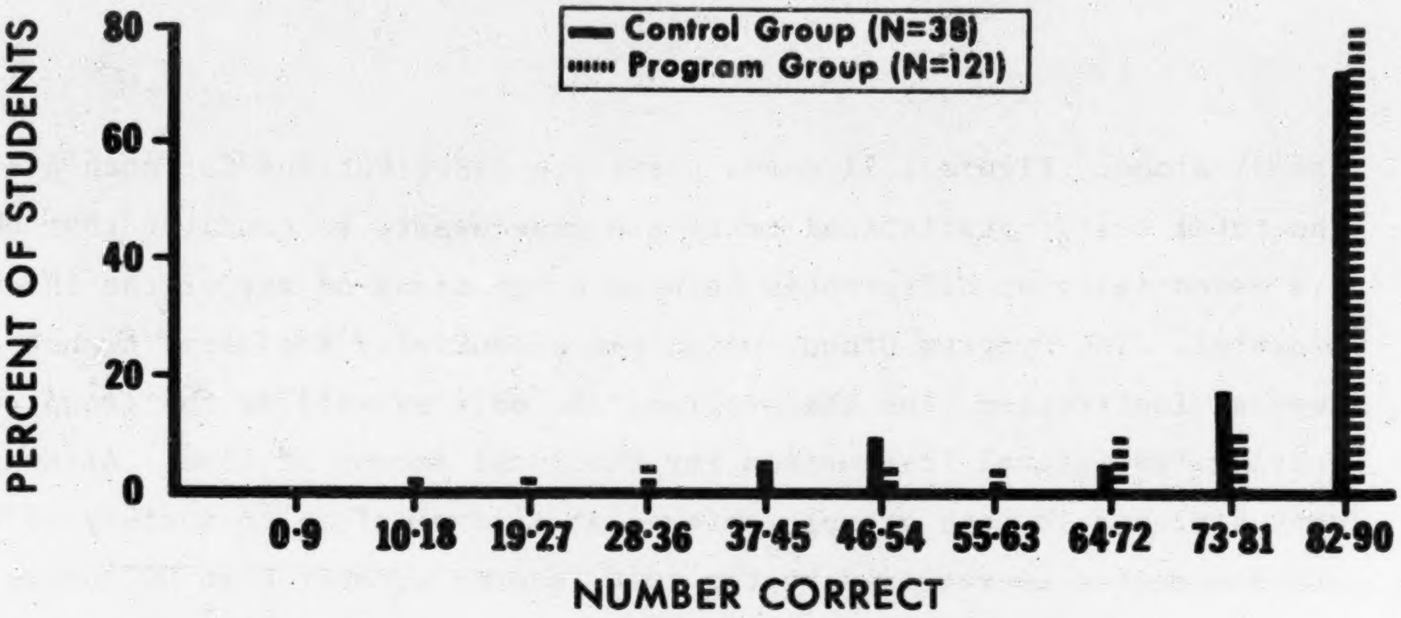

Figure 1.11 Score distributions of control and program groups on addition and subtraction test.

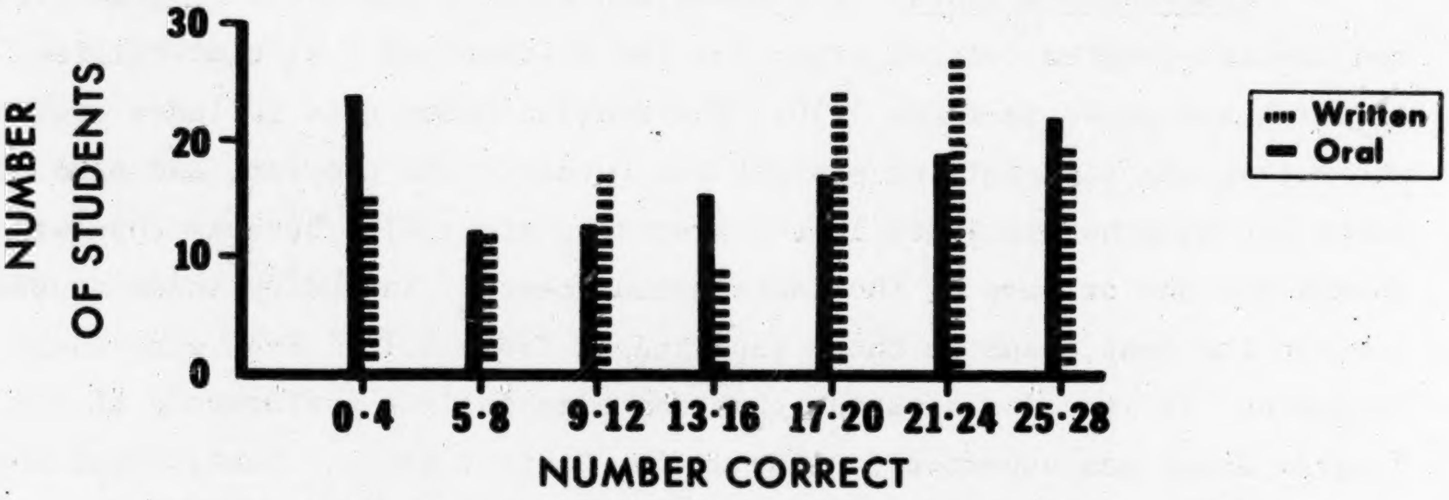

Figure 1.12 Score distributions of program group on end-of-year time-telling test (N=121).

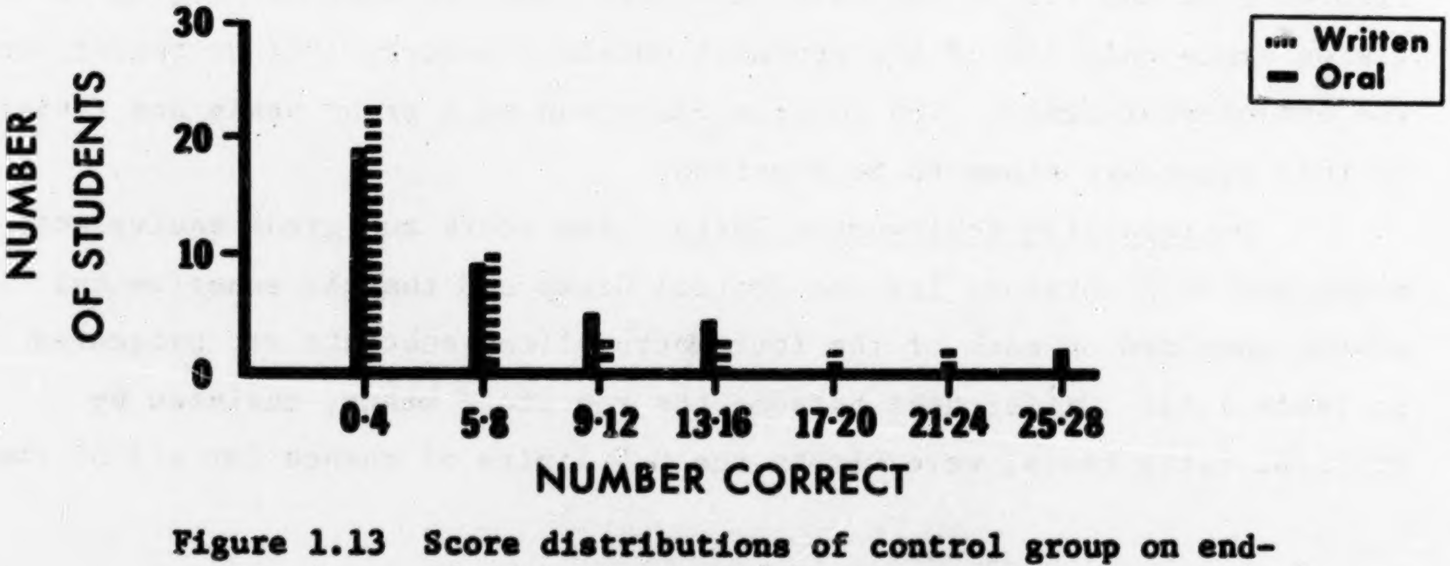

Figure 1.13 Score distributions of control group on end-of-year time-telling test (N=38).

subtests (CR's ranging from .29 to 1.89). The results indicate that the achievement levels of the program and non-program groups at the end of the first grade were equivalent in the three subject matter areas for which neither received programmed instruction; and were also equivalent in arithmetic achievement for which the experimental group received teacher instruction plus programmed instruction during the school year. (Figure 1.14 shows score distributions for each group on the arithmetic subtest of the Metropolitan Achievement battery.) Thus, using the program in addition to the teacher resulted in no additional gain, and the extra time spent on the program did not affect achievement in other subjects.

Summary

End-of-the-year testing in the first grade and comparisons of the experimental classes with the control classes appeared to indicate the following: (1) When both programmed instructional material and the teacher are employed to teach the same thing, there is a redundancy of effort so that no additional student achievement is attained. However, the effects of the program must be known so that the decision can be made as to where the teacher needs to supplement the program and where the teacher can assume that certain knowledges have been taught so that more advanced learning can proceed. (2) When a program is used to teach a topic usually not taught intensively by the teacher, it seems possible that the subject taught can be learned and retained. (3) Additional time taken from the usual first grade curriculum for the use of programs in arithmetic topics did not result in decreased achievement in other areas as measured by the Metropolitan Achievement Test.

Table 1.11

Raw Score and Grade Equivalent Means and Standard Deviations for Program and Control Groups on the Metropolitan Achievement Tests

Group		N	Metropolitan Sub-Tests							
			Word Knowledge Total Possible (35)		Word Discrimination Total Possible (35)		Reading Total Possible (45)		Arithmetic Total Possible (63)	
			$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
Raw Scores	Program	121	26.98	6.26	28.76	6.17	30.54	9.07	55.28	5.92
	Control	38	28.34	5.28	29.60	4.44	30.95	7.06	57.34	5.85
Grade Equivalent	Program	121	2.18	.52	2.47	.64	2.40	.72	2.50	.58
	Control	38	2.26	.45	2.49	.52	2.31	.50	2.85	.61

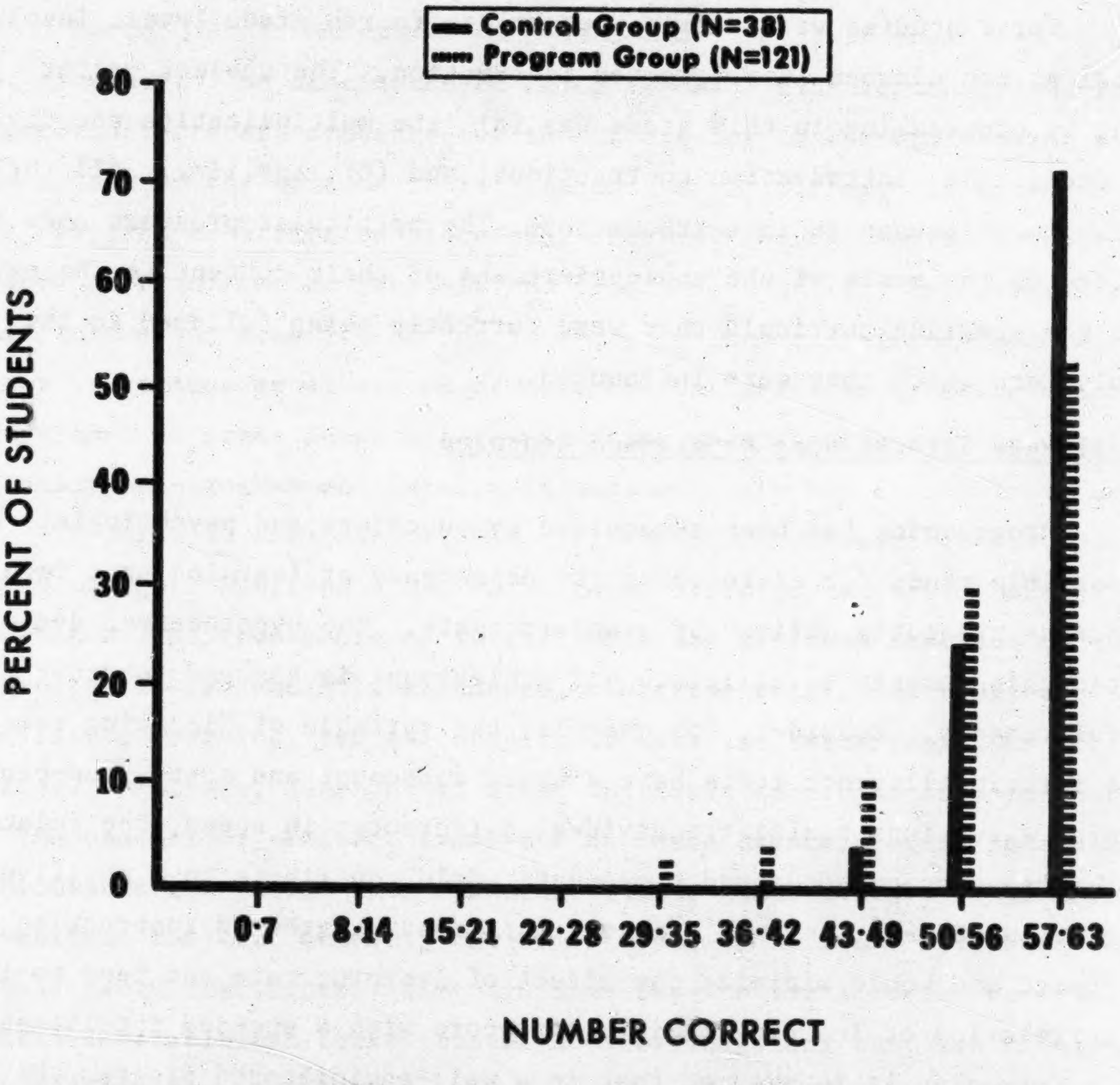

Figure 1.14 Score distributions of program and control groups on arithmetic subtest of Metropolitan Achievement Test.

Arithmetic and Spelling in Grade 4

Overview

Three studies were carried out at the fourth grade level, involving a total of ten classes in programmed instruction. The subject matter taught by programming in this grade was (a) the multiplication and division facts, (b) introduction to fractions, and (c) spelling. All three programs were presented in textbook form. The particular programs used were selected on the basis of the appropriateness of their content to the arithmetic and spelling curricula that were currently being followed in the schools into which they were introduced.

Intelligence Effects upon Programmed Learning

Programming has been recognized by educators and psychologists as one possible means for diminishing the dependency of learning upon intelligence as presently defined by standard tests. The hypothesized decreased relationship between intelligence and achievement is assumed to occur for several reasons. Consider, for example, the variable of "learning rate." Since most intelligence tests have a speed component and most group-paced learning situations maximize individual differences in speed, the relationship between the common speed components would contribute to a positive correlation between the two. In contrast, most programmed instruction is self-paced and would minimize the effect of learning rate and tend to lower the correlation of learning achievement score with a speeded intelligence test. Secondly, it is assumed that in a well-administered program the students have mastered the prerequisite learning for taking the program. This should have the effect of reducing individual differences and lowering the correlation coefficient. Thirdly, if a program is an effective instructional procedure so that more students achieve mastery (top-test scores) than with other instructional procedures, the range of scores is reduced and consequently the size of the correlation coefficient is decreased.

Investigations seeking to determine the raltionships between intelligence test scores and the learning which occurs from a program have yielded varying results, with certain studies (Silberman, 1962) indicating no relationship between IQ measures and achievement while others (Shay, 1961; Porter, 1961; Smith & Moore, 1961) have found at least a moderate relationship. The discrepancies among research findings regarding intelligence effects are probably due to differences in the programming situations used in the various investigations, e.g., differences in subject matter, program construction procedure, program format, length of instruction, manner of presentation, and student population, as well as differences in definitions and measures of intelligence. Since differences such as these occur in various applications of experimental learning situations in the classroom, it seems logical to expect that consistent results regarding the intelligence-programmed learning relationship are not easily obtainable.

Detailed analysis of such differences requires systematic controlled study of the dimensions along which these differences vary. These dimensions consist of the components of intelligence and aptitude that are related to learning, as prerequisite knowledge, the presence of "learning sets" which facilitate learning, and the ability to make the discriminations, e.g., spatial, abstract, linguistic, etc., necessary to profit from instruction. In the absence of adequate knowledge of these dimensions, no generalizable conclusions can result from disparate single experiments. Simple empirical investigations can, however, obtain information on gross variables such as intelligence test scores which are sensitive to gross differences in experimental materials and operations without establishing conclusive evidence for all situations. The accumulations of this type of research can accomplish the following: first, provide information concerning the variable in question to potential users of the particular kind of material and conditions that were investigated; and second, if data from a number of investigations performed under specifically described conditions are obtained, certain general characteristics regarding the effects of certain conditions upon the variable can be described and subjected to more detailed experimental analysis.

The first study in the fourth grade (4A) was of this empirical type, with the objective of evaluating the extent to which learning from an instructional procedure utilizing both programming and teacher presentation was influenced by the intelligence of the learners. The subject matter, taught in daily sessions over an 11-week period, was multiplication and division facts. Students in six classes worked in the program on four days of each week with each student being permitted to advance through the program at his own pace. On the fifth day, the teacher directed a general review and discussion of the material presented in the program. A correlation between intelligence and terminal test performance was obtained to assess the degree of relationship between IQ and achievement. Also, the achievement test performances of high and low intelligence sub-groups were compared.

Review and Acceleration

The same six classes which participated in Study 4A were employed in investigating the use of programmed instruction as a review technique, and also as a technique for accelerating the fourth grade arithmetic curriculum (Study 4B). Following the multiplication and division program, four of the six classes continued with the usual school curriculum for the remainder of the school year. Within this group, two of the classes received no further programmed instruction. The other two were given a retention test on the multiplication and division facts two months following the completion of the program, and individual students not passing this test were required to use the program as an aid in reviewing and relearning the multiplication and division material they had forgotten. The remaining two classes, constituting an accelerated group, began the fractions program (containing material not ordinarily presented at this grade level) soon after completion of the multiplication and division program, rather than continuing with the usual Grade 4 curriculum. Thus by the end of the year these latter two classes had been exposed to intensive instruction on a topic in which those following the usual curriculum had had only periodic and incidental instruction. Achievement measures for all classes were compared at the end of the school year to determine the relative effects of the curriculum, the existing curriculum plus individual programmed review, and the accelerated curriculum, upon the amount of arithmetic knowledge retained at the end of Grade 4.

Effects of Classroom Surroundings

In Study 4C, a spelling program was used to evaluate the effect that variations in classroom environmental stimuli have upon learning. It is an increasingly common practice in present-day school systems to teach various subjects in classrooms specifically designed for instruction in certain subject matter areas, e.g., science rooms, language rooms, etc. The major reason for this practice is, of course, that special equipment too bulky or expensive to be put in the ordinary classroom is necessary for teaching certain subjects effectively. It may be, however, that other learning advantages, more subtle than the use of special equipment, accrue from the practice of using a specific room to teach only one subject matter. The presence of classroom stimuli such as charts or projects completed, particular seating arrangements, and general "classroom atmosphere" are often assumed by educators to have facilitating effects upon learning and retention. A room in which only one subject matter is taught usually has a variety of such stimuli, all pertaining to the single knowledge area. It is possible that the uniqueness of the surroundings, in itself, may facilitate the learning which takes place in such a room.

This assumption is supported to some extent by certain laboratory investigations of human learning which have demonstrated that certain stimuli which are present in the total learning situation enhance learning, while other stimuli interfere with learning (Slamecka & Ceraso, 1960). One theoretical explanation which accounts for these empirical findings is the concept of interference, which has received considerable research attention (Postman, 1961). A basic assumption in many theories of learning is that new responses being learned become attached to all stimuli present during learning, including incidental as well as critical stimuli. According to the interference concept, if the incidental stimuli present are already associated with responses other than the one being learned, these older associations tend to interfere with the new one being made. This associative interference hypothesis suggests that when several different subjects are being taught in the same classroom, the response appropriate to one subject may be associated with classroom stimuli in a way which interferes with learning responses to other subjects. Therefore, a possible advantage of using different classrooms for different learning materials is that the unique surrounding stimuli present in the different classrooms are associated with only one subject matter, which may diminish interference effects upon learning. Also, once a response has become associated with these incidental stimuli, the presence of such stimuli may facilitate retention of material learned if a test is given in their presence, since they would tend to evoke the appropriate associative responses rather than interfering responses.

Although this hypothetical explanation of possible effects of incidental stimuli upon learning is quite crude and general from a rigorous point of view, it does represent a possible extrapolation of existing theoretical descriptions of the learning process to educational practice, and suggests an initial experimental study. Consequently, Study 4C was performed to explore the hypothesis suggested by a general interpretation of the concept of interference, namely, that incidental classroom stimuli will facilitate or inhibit learning of specific responses in accordance with whether or not the responses already associated with them interfere with the new responses required. Specifically, it was predicted that a group receiving both spelling instruction and spelling tests in a special room in which no other subject matter was taught would demonstrate more learning than a group given the same instruction and testing in a room used by that group for learning other subjects as well. A third group, receiving spelling instruction in a special room but all spelling tests in the usual classroom, was used to determine the effect of surrounding stimuli upon test performance alone. It was predicted that test performance of the latter group would be lower than that of the group which received all tests in the special room.

Study 4A:
The Effect of Intelligence upon Programmed Learning Achievement

Method

Subjects. The same six classes in the fourth grade were used for both Studies 4A and 4B. The mean IQ of the 173 students in the combined classes who completed the study was 116.45 (s=10.91).

Design and Procedure. In the first three weeks of school all six of the classes were given a general review of the arithmetic material taught in the previous year (e.g., addition, subtraction, measuring, using large numbers). At the end of the third week, a multiplication and division test, Part A (MD-A) was administered as a pretest to assess level of achievement in multiplication and division operations prior to beginning the multiplication and division program. All six classes were then given the program, under identical conditions, over the next six weeks. In a typical week, a student worked at his own pace through the program during 45-minute work sessions on Monday, Tuesday, Thursday, and Friday. On any of these programmed instruction days, if the student completed one of the ten blocks of material into which the total program was divided he was given a written test on that block before continuing in the program. If a student failed to achieve a score of at least 70% on any block test, he was required to work through that block of the program again and pass a re-test at the 70% level before he could advance to the next block. This procedure provided a check on individual progress, and assured the teacher that students were learning the required material.

The Wednesday sessions were teacher-instruction periods, during which the teacher presented review and practice materials relevant to the parts of the program which most students had completed. The specific content of these teacher sessions was specified in detail in the Teacher's Manual that was constructed for the study, insuring that all classes received the same treatment on these days.

As a result of the self-pacing procedure, some students finished the program before the end of the allotted six-week period, while others were unable to complete it within the designated time limit. Students finishing early were provided with enrichment materials, to be used individually during the sessions when others were still working in the program. These materials consisted of workbook exercises pertaining to the multiplication and division facts presented in the program, as well as problems and activities which required applications of these facts. The slowest students in each class, for whom it was obvious that not enough class time was allotted for program completion, were required to take the program home several night during the last two weeks and work on it there as well as in the classroom program sessions.

At the end of the six-week period, students in all classes were given the two posttests to evaluate the learning which resulted from the self-paced, combined program-and-teacher instructional procedure. One posttest (MD-A) was identical to the pretest, containing 25 of the multiplication and division items in the arithmetic sub-test of the Stanford Achievement Battery, plus 15 similar items constructed by the experimenters and added to give the test a wider range. The second posttest (MD-B) was a 40-item sampling of the items used in the ten block tests. (The MD-A test included some rather difficult items with larger numbers than were used in the program, while MD-B contains simpler facts plus some word problems. In general, MD-A was a more difficult test.)

Results

It was necessary to eliminate six students from the original group before performing the final analyses because of imcomplete data due to absence on critical test days. The pre- and posttest data for all six groups combined are shown in Figure 1.15 for MD-A. For the 173 students for whom complete data was obtained, the correlation between intelligence scores and MD-A posttest scores was .19, and the correlation of IQ and MD-B posttest was .20. Taken alone, these low correlation values suggest that intelligence differences accounted for very little of the variance associated with

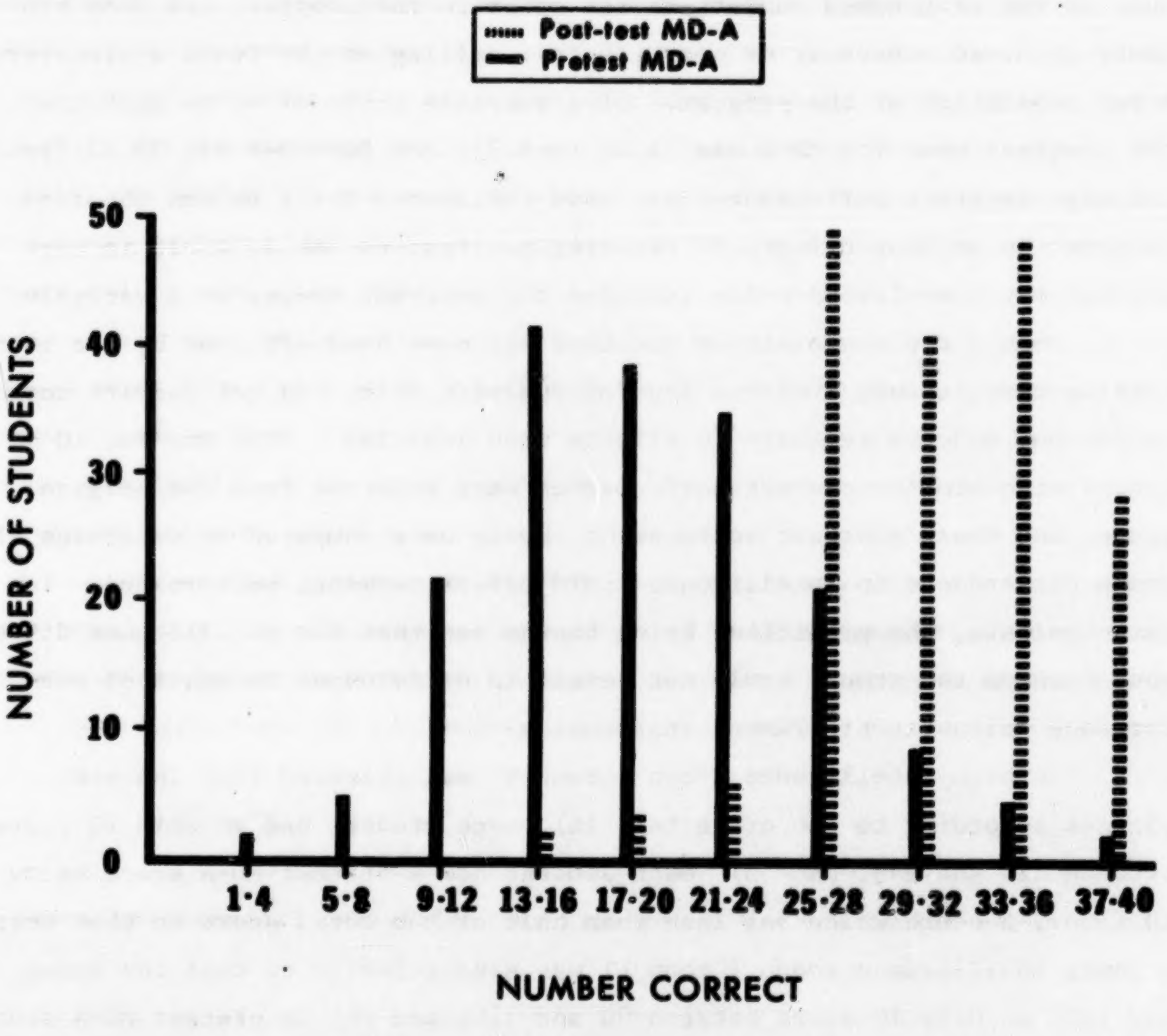

Figure 1.15 Pretest and posttest distributions of scores on test MD-A (N=173).

achievement performance following programmed instruction. However, inspection of the MD-A pretest scores indicated that some students already knew much of the programmed subject-matter prior to the program, and many students obtained scores at or near the test ceiling on the tests administered after completion of the program. Of a possible score of 40 on each test, the posttest mean for MD-A was 31.31 ($s=4.75$) and for MD-B was 33.72 ($s=4.83$). The high posttest performances may have influenced the $\underline{r}$ values obtained because the ceiling effect, by reducing posttest variance, would in turn depress any correlation which included the posttest scores as a variable.

Since the correlations obtained may have been affected by the test ceiling restriction, a second type of analysis which did not require correlation was made to evaluate IQ effects upon learning. High and low IQ groups with similar pretest performances were selected from the original group, and their posttest achievement levels were compared to determine if group differences in intelligence would affect terminal performance. In this analysis, the prediction being tested was that the intelligence difference between the groups would not result in differences in posttest performance following programmed instruction.

A high intelligence group (Group H) was selected from the six classes according to two criteria: (a) each student had an Otis IQ score between 120 and 140, and (b) each student had a pretest MD-A score below 20 (i.e., a score which was less than half of the total score on that test). A lower intelligence group (Group L) was also selected so that the group had (a) an Otis IQ score between 90 and 110, and (b) a pretest MD-A score below 20. The requirement that the pretest score be less than half of the total possible score on the MD-A test insured that all students in both of the groups selected did not have adequate knowledge of the multiplication and division facts prior to taking the program, and also that all students had a considerable test range (at least 20 points) over which improvement could be demonstrated on the posttest. Group H was then compared with Group L on the MD-A posttest to determine whether IQ differences, as determined by the Otis test, resulted in significantly different posttest performances following completion of the program.

Table 1.12 presents the means and SD's for the intelligence, pretest, and posttest scores of the students meeting the criteria for inclusion in Group H or L. Mean gains from pretest to posttest are also shown. Since the difference between the pretest means for the 28 students in Group H and the 27 students in Group L was not significant ($t=1.74$, df/53, $p > .05$), the groups were tested as samples from different intelligence populations which had the same knowledge of multiplication and division prior to receiving the program. Comparison of the differences in posttest means and in mean gains yielded *t* values of 1.25 and .04 respectively, neither of which is large enough to reject the null hypothesis that the groups were from the same or equal achievement populations following programmed instruction. The large difference between the groups in intelligence apparently had no effect upon final achievement performance or the amount of learning which took place in the course of programmed instruction. The results of this comparison and the correlation analysis are consistent, indicating that intelligence as measured by the Otis test had little or no relationship to the learning which occurred under the programmed instruction conditions used in the present study.

Table 1.12

Means and Standard Deviations for High and Low Intelligence Groups on Measures of IQ and Achievement in Multiplication and Division

				MD-A Test Scores 40 items					
		IQ		Pretest		Posttest		Gain	
Group	N	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
H	28	125.86	4.58	14.57	3.26	30.04	5.34	15.46	5.35
L	27	103.04	5.25	13.11	2.99	28.52	3.48	15.41	4.22

Study 4B:
The Use of Programmed Instruction for Review and Acceleration of Arithmetic Learning

Method

Design. After completion of the multiplication and division program, the six classes which had participated in Study 4A were divided into three groups of two classes each and used in Study 4B. Two of the three groups were presented the arithmetic curriculum that was currently being followed by the school system, using teacher instruction for the remainder of the year. In order to guarantee that all four classes in both of these groups were exposed to the same subject matter at the same time, the teachers followed weekly lesson plans which were outlined in a Teacher's Manual specially constructed for this study. The plans indicated what topics were to be covered each week, and the textbook and workbook units and pages to be used in presenting them. Within this framework, the only difference in treatment was that students in one group (Group R) received a programmed review of multiplication and division one month following the program, while the second group (Group NR) did not. The third group (Group F) was given programmed instruction in fractions soon after completion of the multiplication and division program. This latter treatment constituted an acceleration of the arithmetic curriculum, since intensive instruction in fractions was normally not a part of the fourth-grade subject matter at the schools participating in the experiment. At the end of the school year the three groups, and a control group which received only traditional instruction throughout the fourth grade, were given a series of tests to determine the effects of acceleration and review upon arithmetic achievement.

Subjects. An attempt was made to organize the intact classes participating into experimental and control groups which were equivalent on three independent measures obtained prior to the study: Otis intelligence test scores, scores on the arithmetic subtest of the Stanford Achievement Tests, and battery median scores on the Stanford Tests. The latter two measures were presumed to be indicators of prior knowledge in arithmetic and general academic achievement, respectively, both of which could be expected to have at least a moderate influence upon the learning required in the study. Table 1.13 gives the means and standard deviations for each group on these measures, after eliminating all students who failed to complete the study because of moving or extended absence. The means of the groups are equivalent on each of the two pre-experimental achievement measures. It was not possible to arrange all classes into equivalent groups on the intelligence measures also. The IQ mean of Group NR is significantly higher than the means for both the Control Group (C.R.=2.70, $p < .01$) and Group F (C.R.=2.38, $p < .05$), and is beyond the .10 significance level when compared to Group R (C.R=1.79, $p < .10$). All differences in intelligence means between Group R, Group F, and the Control Group are well within chance limits, however, with CR's ranging from .46 to a maximum of 1.09.

Procedure. Groups R and NR received identical treatments with one exception. One month after completion of the multiplication and division program, Group R was readministered the MD-A test (described in the previous study) as a retention test. Any student failing to achieve a score of 80% or higher on the retention test was required to work through the sections of the multiplication and division program which taught the materials he had either forgotten or else had not learned previously. This use of the program as a review and relearning tool took place at home by the individual students requiring it, while the group continued with the on-going curriculum in the regular daily arithmetic class periods. During this time Group NR received no retention test or review, but simply continued with the teacher-directed curriculum normally employed at the school.

Group F, receiving an accelerated curriculum with programmed instruction, began working on the fractions program three weeks after completing the multiplication and division program. (A vacation, followed by a period of teacher instruction on two-figure multipliers, occurred during the three-week interval between programs.) After administration of the 50-item pretest to determine existing knowledge of fractions, the fractions program was administered over an eight-week period by the same general procedure as was used for the multiplication and division program.

Table 1.13

Group Size, and Means and Standard Deviations of the Grade Four Arithmetic Groups on Measures of IQ, Arithmetic Achievement, and General Academic Achievement Taken Prior to Beginning Study 4B

Group	N	IQ*		Arithmetic Achievement**		Academic Achievement***	
		$\overline{X}$	s	$\overline{X}$	s	$\overline{X}$	s
F	56	119.55	12.69	4.72	.77	5.03	.82
R	65	115.48	9.23	4.97	.80	4.91	.91
NR	52	114.33	10.00	4.76	.71	4.93	.81
Control	59	113.39	11.70	4.79	.83	5.15	.94

*Otis Quick-Scoring Mental Ability Tests (alpha)

**Arithmetic subtest of Stanford Achievement Battery (grade-placement scores)

***Battery Median of Stanford Achievement Battery (grade-placement scores)

On four days of each week, students worked in the program on a self-paced basis, taking block tests as necessary and repeating the block if test performance was below criterion. Arithmetic sessions on the fifth day of each week were devoted to practice and review under the teacher's direction, using small groups that were relatively homogeneous in their rates of progress through the program.

Upon completing the fractions program, Group F received teacher instruction in other arithmetic topics usually taught in the fourth grade for the remainder of the year. In this phase, however, the rate of progress scheduled in the Teacher's Manual was considerably faster than the rates of Groups R and NR, with the consequence that all three groups ended the school year at the same point in the text being used. Thus Group F had covered (in a shorter time) all major arithmetic materials covered by R and NR, and in addition had been exposed to eight weeks of intensive instruction in an advanced topic. The total time spent on arithmetic was equal for the three groups.

At the end of the year, all three groups were given the following tests: the two multiplication and division tests, MD-A and MD-B; the fractions test; and the Stanford Achievement Battery. The Control Group, which received no programming at all during the year, was also given all of the tests.

Results

The three experimental groups had each completed the multiplication and division program, and taken the MD-A and MD-B tests, just prior to the beginning of the acceleration study. Since the intelligence mean of Group NR was higher than the other two groups, the three groups were first compared on these MD-A and MD-B tests to determine if intelligence differences were producing a systematic effect upon the group performances at the start of Study 4B. Table 1.14 gives the means and SD's of each group on the MD-A and MD-B measures that were administered prior to the beginning of the study. An analysis of variance indicated no significant differences between the groups on the MD-B test ($F < 1.00$, df/2, 170, $p > .05$), but a significant difference among the groups was found on the MD-A test ($F = 3.52$, df/2, 170, $p < .05$), due to the low mean of Group NR. Although having the higher intelligence score, the NR group was significantly lower than Group F ($t = 2.78$, df/106, $p < .01$) on the achievement measure taken following administration of the multiplication and division program. These data add support to the conclusion reached in Study 4A, that IQ had little effect upon programmed learning, and suggest that the IQ differences among the groups may be ignored in analyzing the results of Study 4B.

The results of the five tests administered at the end of the school year are presented in Table 1.15. Variance analyses of the Stanford Battery

Median and the Stanford Arithmetic subtest results showed that differences among the four groups were not significant for either measure (F=1.37 and 2.23, respectively, df/3, 228, $p > .05$). The equivalence of all groups on these standardized tests indicates that they had advanced to equal degrees during the school year in terms of general academic achievement, and also in number reasoning and general arithmetic facility as measured by the Stanford subtests. There is no indication that the groups were not comparable in the ability to learn.

Table 1.14

Means and Standard Deviations for Groups F, R, and NR on the MD-A and MD-B Tests Administered Upon Completion of the Multiplication and Division Program

Group	N	MD-A Total Possible-40		MD-B Total Possible-40	
		$\overline{X}$	s	$\overline{X}$	s
F	52	32.40	4.77	33.94	4.51
R	65	31.54	5.10	33.85	5.28
NR	56	30.04	3.95	33.36	4.55

In contrast to the equivalence found among groups on the standardized tests, reliable differences were obtained when the more specific tests evaluating the programmed materials were analyzed. The analysis of variance for the MD-A test was significant (F=3.53, df/3, 228, $p < .05$), with further *t* tests showing that the Control Group mean was significantly lower than the Group F and Group R means beyond the .02 level (t=2.83 and 2.50, respectively). For the MD-B test, it is apparent from Table 1.15 that the variance of the Control Group is much larger than the variances of all the experimental groups. Also, the Control Group mean on the MD-B test was significantly lower than the F, R, and NR groups (t= 3.74, 5.62, and 4.70, respectively), at levels beyond p=.01 in every case. Almost without exception, then, the end-of-year performance of the Control Group was inferior to all of the experimental conditions on the multiplication and division tests. However, no significant differences among the three experimental groups were found on these tests. The latter finding fails to support the expectation that the use of the program as a review and relearning tool in Group R treatment would facilitate end-of-year achievement in multiplication and division. This finding may in part be dictated by the fact that the mean of each group was very near the test ceiling and that the test did not account for sufficient variance among individuals.

Analyses of the fractions test data showed that Group F performance was superior to all other groups. The t values obtained in comparing the Group F mean with Groups NR, R, and the Controls were 6.88, 6.83, and 6.06, respectively, each of which is well beyond the .001 level of significance. The lower SD of Group F relative to the others indicates that its superior achievement in this topic was quite consistent among the students who took the fractions program. Further t tests gave no evidence of significant differences among the NR, R, and Control groups. The score distributions on the fractions pre- and posttests are shown in Figure 1.16. The N in this distribution is 45, since fractions pretests were available for only 45 of the 52 students in Group F. Figure 1.17 shows the fractions posttest for the R and NR Groups combined.

The finding that the group receiving a program in fractions performed better in fractions than those who did not is not surprising in itself. However, Group F spent eight weeks of arithmetic periods learning the additional advanced material, necessarily taking away learning time from the usual fourth-grade arithmetic activities in which the remaining groups were engaged, and yet gave no evidence at the end of the year that it had learned or retained less of the other arithmetic topics ordinarily

taught in Grade 4. The fact strongly suggests that acceleration of the curriculum produced a significant amount of additional arithmetic learning without being detrimental to the learning and retention of the material in the standard curriculum.

Table 1.15

Means and Standard Deviations of Groups F, R, and NR and the Control Group on All Arithmetic Tests Administered at the End of the School Year

Group	N	Stanford Battery Median*		Stanford Arithmetic*		MD-A Total Possible-40		MD-B Total Possible-40		Fractions Total Possible-50	
		$\overline{X}$	s	$\overline{X}$	s	$\overline{X}$	s	$\overline{X}$	s	$\overline{X}$	s
F	52	6.36	.94	6.46	.97	37.48	3.10	36.40	4.61	43.75	4.85
R	65	6.24	1.03	6.44	1.05	37.06	2.75	37.66	2.30	32.89	10.55
NR	56	6.06	.83	6.29	.90	36.75	3.23	36.98	2.93	33.71	9.39
Control	59	6.06	.87	6.04	.98	35.61	3.68	31.20	8.90	34.48	9.92

*Grade placement

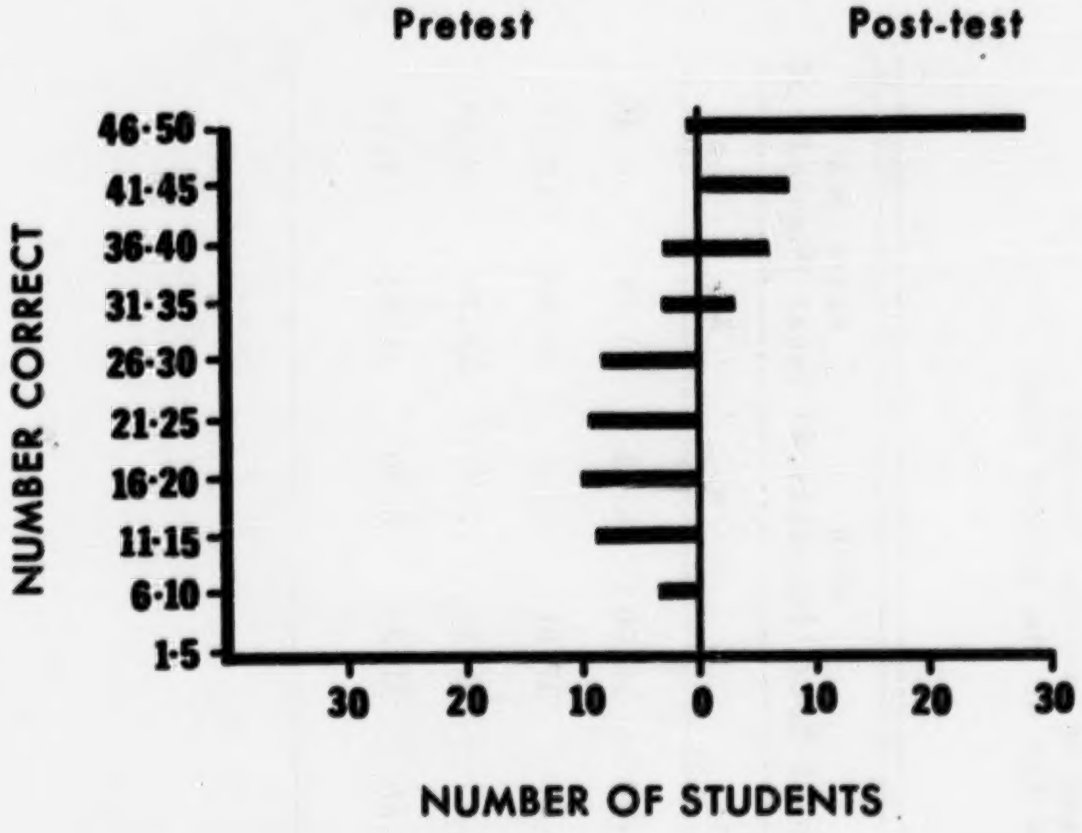

Figure 1.16 Pretest and posttest scores on fractions test, Group F (N=45).

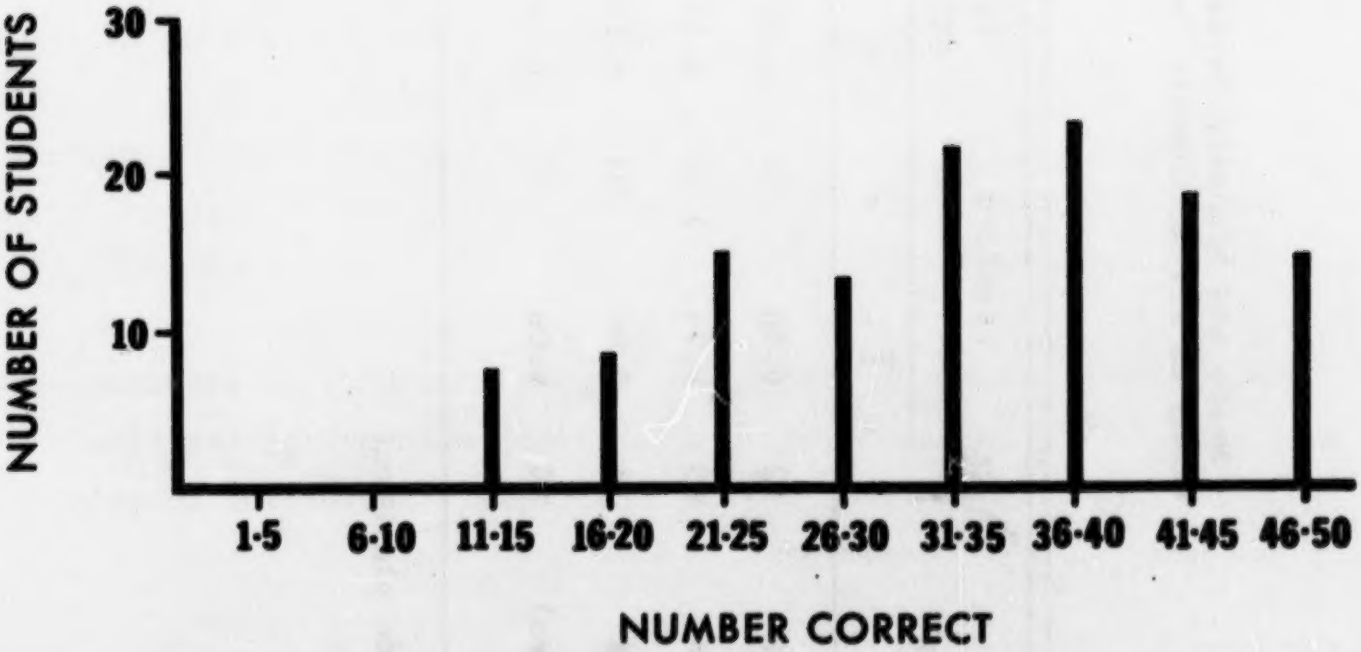

Figure 1.17 Distributions for Groups R and NR (N=121) combined on fractions posttest.

Study 4C:
Effects of Classroom Surroundings Upon Programmed Learning in Spelling

Method

Design. Three groups, taking a full year's programmed spelling course, were used in this study. One group (Group H) received programmed instruction and all weekly spelling tests in teh same homeroom used for all other instruction. A second group (Group S) received all programmed instruction and all weekly spelling tests in a special room, used for no purpose by this group but learning spelling. The third group (Group S-H) received the program in a special room set aside for spelling instruction only, but took all weekly spelling tests in its homeroom. These groups were compared on a series of spelling tests given at the end of the school year to determine if programmed spelling achievement is influenced by variations in the surroundings present during learning and testing. A Control Group, receiving traditional spelling instruction in its own homeroom, was also compared with the experimental groups to evaluate the effectiveness of programmed instruction in teaching fourth grade spelling.

Subjects. Two of the six classes chosen to participate in Study 4C were also participants in the Arithmetic Studies 4A and 4B. Four other classes were selected so that, when combined into two more groups of two classes each, the resulting three groups were equivalent in terms of mean IQ, mean spelling achievement level, (as measured by the spelling subtest of the Stanford Achievement Tests), and mean level of general academic achievement (as measured by Stanford battery median scores). A Control Group, consisting of two classes equivalent to the experimental groups on these same measures, was also selected. It was necessary to eliminate some students (about 4% from each experimental group) when making the final analyses, for reasons such as extended absence or transferring schools during the course of the experiment.[6]

---

[6] For this reason, the size of Group S in Study 4C differs slightly from the same group used in 4A and 4B. Students that had to be eliminated from one study were not necessarily eliminated from all others because of differences in length of the studies, time of year, etc.

Table 1.16 describes the four groups used in the final analyses, listing their sizes and the means and SD's of the independent variables on which they were matched. As was the case in Studies 4A and 4B, the three groups in 4C were bright-normal in intelligence, and grade placements on the Stanford spelling subtest and the total Stanford battery indicated that their achievement levels were considerably above average at the beginning of the fourth grade. Statistical tests for differences between groups on the various independent measures did not reveal any significant differences, so the groups were assumed to be equivalent on all measures taken at the beginning of the school year.

Procedure. In the first week of school the three experimental groups were given additional pretests to assess level of spelling achievement prior to beginning the study. The Program Pretest consisted of 60 words randomly chosen from the 354 new words presented in the program that was used. This test required the students to write the words as they were pronounced by the teacher. The other pretest given was the Spelling subtest of the Iowa Tests of Basic Skills (Form 1).[7] The Iowa pretest contained 38 multiple-choice items, each of which required the student to indicate on a special answer sheet which of the four printed words was spelled incorrectly.

Following the pretests, the experimental groups began a 29-week period (out of a total of 37 school weeks) of programmed instruction in spelling. The general procedure for all groups was to work on assigned frames during scheduled 20-minute class periods on Monday, Tuesday, and Wednesday of each week. On Thursday the teacher directed a review and enrichment period which provided further practice with the words presented in the week's assignment, and on each Friday students received a written test of the words covered during the week. Exceptions to this procedure occurred on five occasions, usually close to a vacation period, at which times the weekly tests were postponed until the next school week. In all, each group received 24 weekly tests.

---

[7] Iowa Tests of Basic Skills. Boston: Houghton-Mifflin, 1955.

Weekly assignments were specified for all teachers in a Teacher's Manual specially constructed for the study. The assignments averaged about 120 frames, an amount which made it necessary for students to complete approximately 40 frames during each 20-minute program work period. At the beginning of each week students were told the frame numbers which constituted the week's assignment, and then were permitted to work at their own paces. Faster students who finished in less than the three 20-minute periods allotted were given individual spelling enrichment tasks by the teacher. Those in danger of not finishing an assignment in the allotted three-day program period were given extra time to insure that they would be able to participate in the teacher-directed enrichment period which followed on the fourth day. Suggested activities for this latter period were given in the Teacher's Manual, and teachers were asked to follow these suggestions to minimize differences among the groups.

The number of words presented per week in the program varied from 16 to 20, in accordance with the length of the assignment made. Some words were repeated on several different weeks because the program contained periodic word reviews as well as new words throughout.

Within this general framework, the experimental conditions for evaluating the effects of classroom stimuli upon learning were arranged as follows: Group H received all instruction and testing in the same classroom in which it received all other instruction (at this grade level classes were self-contained). Group S-H received programmed instruction and teacher-directed enrichment in a separate room, used only for learning spelling. All weekly tests and the final posttests were administered in the homerooms of these Group S-H classes, however. Finally Group S received all instruction, and all tests as well, in a special spelling room which it used for no other purpose. (The special rooms used by the latter groups contained no displays, exhibits, or other material associated with any topic other than spelling.) The Control Group followed the usual spelling curriculum in the regular classroom, using the standard curriculum. This curriculum contained 112 of the 354 new words presented in the program, plus other words which were generally less difficult than those on the programmed spelling list.

Table 1.16

Group Size, and Means and Standard Deviations of the Grade Four Spelling Groups on Measures of IQ, Reading Achievement, and General Academic Achievement Taken Prior to Beginning Study 4C

Group	N	IQ*		Spelling Achievement**		General Achievement***	
		$\overline{X}$	s	$\overline{X}$	s	$\overline{X}$	s
H	47	114.17	14.62	5.07	1.08	4.77	.88
S-H	67	115.78	9.18	5.01	.92	4.89	.90
S	55	114.95	11.35	5.21	.98	4.86	.89
Control	59	113.39	11.70	5.45	.88	5.16	.94

*Otis Quick-Scoring Mental Ability Tests (alpha)

**Reading subtest for Stanford Achievement Battery (grade-placement score)

***Battery Median of Stanford Achievement Battery (grade-placement score)

Upon completion of the program, the three experimental groups and the Control Group received three posttests assessing spelling achievement. The tests used were the same Program and Iowa tests administered to the experimental groups prior to beginning the study, and the Spelling subtest of the Stanford Achievement Battery.

Results

Pretests. Means and SD's of the two pretests, which were administered to the three experimental groups to assess their equivalence in pre-experimental spelling achievement, are presented in Table 1.17. Critical

Table 1.17

Means and Standard Deviations for Groups H, S-H, and S on the Iowa and Program Spelling Pretests

Group	N	Iowa Pretest (38 items)		Program Pretest (60 items)	
		$\overline{X}$	s	$\overline{X}$	s
H	47	20.70	7.86	31.28	14.07
S-H	55	19.67	8.56	35.51	11.38
S	67	20.64	7.01	34.46	10.88

ratio tests showed that mean differences among all groups were within the .05 limits of chance on both pretests, indicating that the groups were equivalent in spelling performance at the beginning of the study.

Posttests. Table 1.18 contains the data for all of the spelling posttests that were administered to Groups H, S-H, S, and Controls at the end of the school year. Figure 1.18 shows pre- and posttests for the program groups and Figure 1.19 shows posttests for the Control groups. Analyses of variance were performed for each of the three tests to determine if differences among the groups were reliable. The F values obtained for the Stanford and Iowa posttests were both less than 1.00, indicating that all groups were equivalent in spelling performance on these two measures. On the Program posttest, however, an F of 10.26 was obtained, which is significant at the .001 level of probability with df/3, 224. Further t tests between the group means on this test showed that the Control Group performance was significantly lower than that of all three of the groups which had received the program (the t values were 4.46, 3.76, and 6.10 for comparison of the Controls with Groups H, S-H, and S, respectively, with

$p < .001$ in each case). The t values obtained when the program groups were compared with each other were .21, .59, and 1.01, none of which is large enough to reach the .10 level of significance.

Summary

These data indicate without exception (a) that the experimental variations in stimulus surroundings had no differential effects upon end-of-year spelling achievement of the experimental groups; and (b) that the use of a program for spelling instruction resulted in achievement equal to the control classes as measured by two standardized spelling tests, and facilitated spelling performance of the program groups relative to traditional instruction on a test which sampled the specific words taught in the program. Little can be said about the first result, except that the hypothesis that the presence of incidental classroom stimuli affects learning was not supported under the conditions used in this experiment. Concerning the second finding, it seems useful to point out that the Program tests, consisting entirely of words taught in the spelling program, contained many words, 39 out of 60, which students in the Control Group did not encounter in the course of their traditional fourth-grade spelling instruction. While this fact explains the lower Control Group performance on the Program test, it also indicates that the program groups, by being required to do so, were able to learn more spelling during the year than were the controls. The additional material apparently did not prevent the program groups from learning the words usually required at this grade level, since students receiving the program performed as well as control students on the standardized tests. Had a more comprehensive evaluation been made, including all of the words to which each group was exposed during the entire year's spelling instruction, more definite information would have been obtained concerning the spelling efficiency resulting from the program and control conditions. Unfortunately, the amount of class time required for such an evaluation made it prohibitive in this study. The present results suggest, however, that further research involving more comprehensive testing would be fruitful in indicating the quantity of spelling material students at this grade level can retain following exposure by programming and traditional methods.

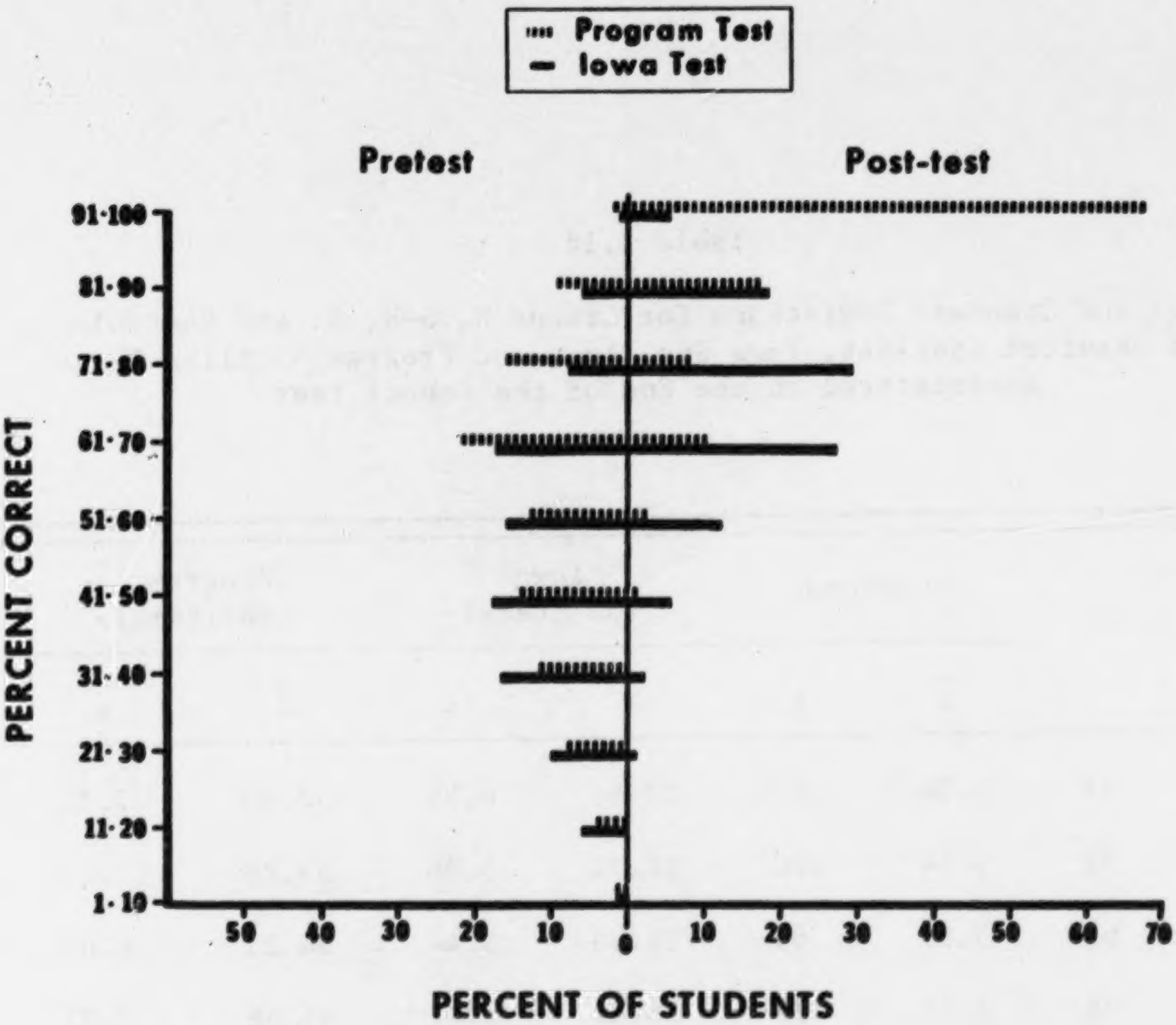

Figure 1.18 Pretest and posttest scores of program groups on program test and Iowa Spelling Test.

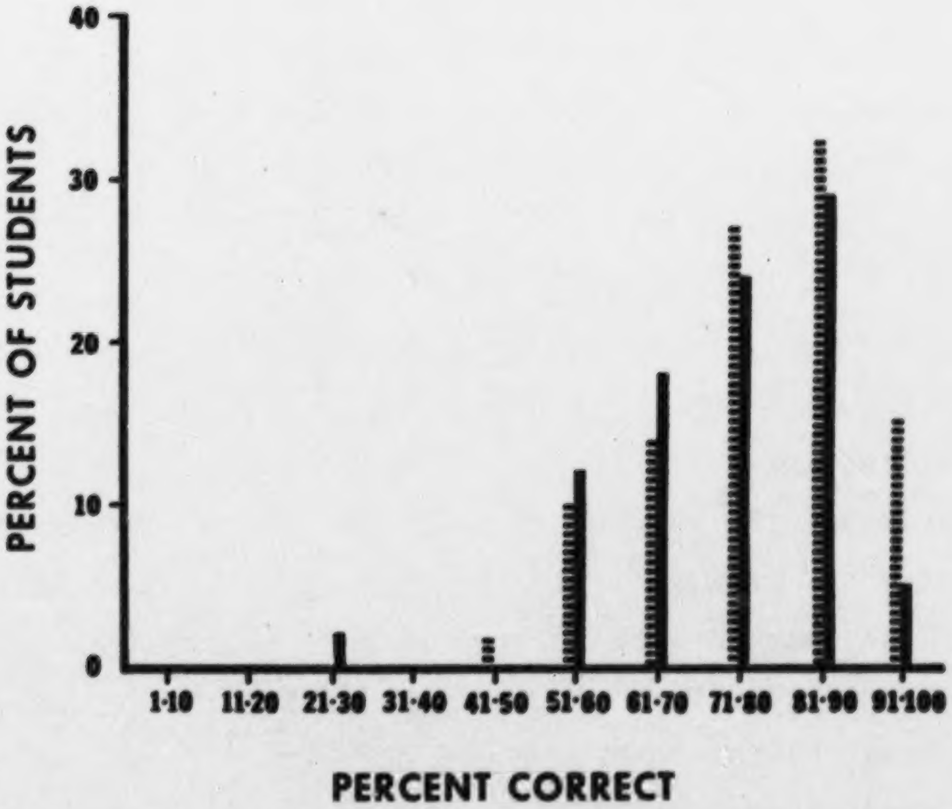

Figure 1.19 Posttest scores of control group on program test and Iowa Spelling Test.

Table 1.18

Means and Standard Deviations for Groups H, S-H, S, and Controls on the Stanford Spelling, Iowa Spelling, and Program Spelling Tests Administered at the End of the School Year

Group	N	Stanford		Iowa (38 items)		Program (60 items)	
		$\overline{X}$	s	$\overline{X}$	s	$\overline{X}$	s
H	47	5.34	.87	27.87	6.59	52.83	7.73
S-H	55	5.64	.90	28.71	6.06	53.24	11.13
S	67	5.61	.84	28.03	5.44	54.21	6.63
Control	59	5.57	.99	28.88	6.02	46.58	7.31

Implications

The specific conclusions of each study have been reported above in the separate studies; the broad implications that can be gathered are the following:

1. There is extensive variation in rate of learning among students when they are given the opportunity to proceed at their own rates with programmed learning materials.

2. Pretest scores show that many of the students know the subject being taught and that some students are not ready to learn it.

3. Different types of teacher-program combinations in several grades made little difference in student achievement.

4. Young children can be taught a subject intensively with little loss in retention (at least over the short time measured in the time-telling study).

5. The extent of the correlation between general intelligence and achievement as a result of programmed instruction depends upon the particular program involved. In general, intelligence appears to be related to the pace with which the student goes through a program.

6. Extension of the curriculum with programmed materials, necessarily taking away from time spent in conventional grade-level instruction, produced additional learning without being detrimental to the learning of materials usually taught at that grade level. In general, students required to learn more did learn more.

Most impressive in these studies was the wide variation in student rate of learning and the wide variation in student achievement prior to instruction. As a result, attention in subsequent academic years at the Baldwin-Whitehall Schools was focused on the individualization of instruction. Some initial pilot studies, undertaken prior to the establishment of an experimental individualized elementary school, are reported in Chapter 4.

References

Berlyne, D. E. *Conflict, arousal, and curiosity*. New York: McGraw-Hill, 1960.

Gilbert, T. F. A structure for a coordinated research and development laboratory. In R. Glaser (Ed.), *Training research and education*. Pittsburgh: University of Pittsburgh Press, 1962. Pp. 559-578.

Glaser, R. Implications of training research for education. In E. R. Hilgard (Ed.), *Theories of learning and instruction*. Sixty-third yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press, 1964. Pp. 153-181.

Glaser, R. Research and development issues in programmed instruction. In R. T. Filep (Ed.), *Prospectives in programming*. New York: Macmillan, 1963. Pp. 278-309.

Glaser, R. & Reynolds, J. H. Instructional objectives and programmed instruction: a case study. In C. M. Lindvall (Ed.), *Defining educational objectives*. Pittsburgh: University of Pittsburgh Press, 1964. Pp. 47-76.

Glaser, R., Reynolds, J. H., & Weinstein, I. P. *Time-Telling: A Teacher-Student Programmed Sequence*. Pittsburgh: University of Pittsburgh, 1963.

Iowa Tests of Basic Skills. Boston: Houghton-Mifflin, 1955.

Joint Committee on Programmed Instruction and Teaching Machines. Criteria for Assessing Programmed Instructional Materials. *Audiovisual Instruction*, February, 1963, *8*, 84-89.

McGeoch, J. A. & Irion, A. L. *The psychology of human learning*. New York: Longmans, Green, & Co., 1952.

Noble, C. E. Comments on the Gibson theory of verbal learning. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior*. New York: McGraw-Hill, 1961. Pp. 217-221.

Porter, D. An application of reinforcement principles to classroom teaching. Cambridge, Mass.: Laboratory for Research in Instruction, Graduate School of Education, Harvard University, 1961.

Postman, L. J. Retention as a function of degree of overlearning. *Science*, 1961, *135*, 666-667.

Reynolds, J. H. & Glaser, R. Repetition and spaced review in the learning of connected discourses. Pittsburgh: Learning Research and Development Center, University of Pittsburgh, 1963.

Shay, C. B. Relationship of intelligence to step size on a teaching machine program. Journal of Ed. Psychology, April, 1961, 52, 98-103.

Silberman, H. F. Self-teaching devices and programmed materials. Rev. educ. res., April 1962, 32, 185.

Slamecka, N. J. & Ceraso, J. Retroactive and proactive inhibition of verbal learning. Psychol. Bull., November, 1960, 57, 449-475.

Smith, W. & Moore, J. W. Size-of-step and achievement in programmed spelling. U. S. Office of Education Report, Grant No. 7-36-087-00. Lewisburg, Pa.: Bucknell University, September, 1961

Underwood, B. J. Ten years of massed practice on distributed practice. Psychol. Rev., 1961, 68, 229-247.

PRECEDING PAGE BLANK-NOT FILMED

CHAPTER 2

THE IMPACT OF PROGRAMMED INSTRUCTION ON SELECTED STUDENT VARIABLES

C. M. Lindvall[1]

Introduction

The studies reported in this chapter are concerned with a variety of student variables, with what changes take place in them and how they relate to one another when students use programmed instruction under certain conditions. The variables studied include student attention, attitude, aptitude, reading ability, speed in working through a program, and selected measures of achievement. The studies will be reported under two major categories:

I. Changes in student attention and attitude over an extended period of use of programmed materials.

II. An analysis of the relationship between achievement and aptitude when various measures are employed under a variety of instructional conditions.

The data presented here must, for the most part, be considered as descriptive data. That is, they provide a description of what happens to certain student variables when programmed instruction is used on a broad scale in rather typical on-going school situations. As a result, it is possible to present data bearing on several important hypotheses; but it is impossible, due to limitations in sampling and in the way in which students could be assigned to various situations, to present rigid tests of these hypotheses. Where tests of statistical significance are provided, it has been largely to provide evidence concerning the internal validity of the results. In other cases only the descriptive statistics have been provided. Despite these limitations, these studies should be of considerable interest and importance since, in general, they are based on more subjects and on a much more extended use of programmed materials than have most previous studies dealing with these important variables. Among other things, the results reported here should be useful in suggesting hypotheses for a more complete study of these measures.

---

[1] With the assistance of Lois Lackner, Joseph Ferderbar, James McCormick, and Glenn Graham.

Studies of Pupil Attention and Pupil Attitude

There is some reason to feel that when a pupil is studying with a good program his degree of attention to the instructional material should be greater than it would be in the typical instructional situation. Contributing to this interest should be the fact that the student is experiencing steady reinforcement as he receives feedback concerning the correctness of his responses and the fact that he is being permitted to progress in the learning activity at his own pace. Of course, the extent to which the use of a program holds the interest of a student is going to depend greatly upon the quality of the specific program, but many researchers in studying the use of programmed materials in relatively limited or laboratory-type situations have reported a high degree of interest among subjects.

On the other hand, many critics claim that asking a student to sit at a desk day after day, turning the pages of a programmed textbook and reading a seemingly endless succession of similar frames, will create a situation that will quickly become extremely boring. Also, some teachers who have used programs in this way report that this boredom does result.

In view of this situation it would seem to be important to investigate this variable of pupil attention and pupil attitude under varying procedures for using programs and at various grade levels. This is the purpose of the studies described in this section of the report. As measures of pupil interest, the studies here employ two procedures: (1) a record of pupil inattentiveness obtained through observation, and (2) measures of pupil attitude obtained through the use of specially constructed attitude scales.

Studies of Pupil Inattentiveness

Studies of overt manifestations of pupil inattentiveness were carried out with three different groups: (1) six classes of first graders using the time-telling program, (2) five classes of fourth graders using the multiplication and division program, and (3) six classes of fourth graders using the spelling program.

A measure of inattentiveness for a class was obtained by having an observer sit in the classroom and, at two-minute intervals, make a quick count of the number of students not giving overt attention to the material they were supposed to be studying (i.e., were gazing about the room, watching another student, day-dreaming, etc.). To obtain a measure for a class at a particular stage in its use of the program, the class was observed for a period of approximately one-half hour on each of the indicated days. This measure used was obtained by determining the mean number of students inattentive at any two-minute time interval over the observation period and converting this to a percent of the total number of students in the class.

Every class that was observed while using programmed materials was also observed, on the same day, while it was studying some type of non-programmed materials. This permitted a comparison of attentiveness under the two conditions. Also, each class was observed on a series of designated days spaced out over the weeks or months that the program was used. This permitted a study of changes in attentiveness over time both in program and non-program classroom situations.

In the following sections summarizing the results when this type of study was carried out with groups using the three different programs, provision is made for comparing inattentiveness under programmed and non-programmed conditions and for noting any trends in inattentiveness over time. Also, data are presented on the interaction between study conditions and time. In each case an analysis of variance has been used to test the significance of the results. It is, of course, recognized that since there was no random selection of programs used or of the classes involved, the generalizations that can be made from any one study are quite limited. However, the significance tests provide some indication of the "internal validity" of the results, and the findings should be useful for suggesting hypotheses for more definitive studies.

Inattentiveness Among First-Grade Students Using a Time-Telling Program

Six first-grade classes in three different schools were observed while they worked through the time-telling program and also while they worked at non-programmed seat work. In studying the program a pupil worked through a separate booklet during each of the 14 periods that he devoted to this material. At any given session, the teacher would first have the class, as a group, read and respond orally to the first few frames. Following this, each pupil would work on his own until he had completed specified sections of the booklet. There were some differences among the six classes in the way in which study sessions were spaced but since a preliminary analysis of the data indicated that these differences were not related to the measures of inattentiveness, the results from all six classes were analyzed together. In every class, 14 separate study periods were spent on the program.

To get a picture of changes in inattentiveness, each class was observed during the first session, the fifth session, the ninth session, and the fourteenth session. The classes were also observed on these same days while they were doing non-programmed seat work.

Tables 2.1 and 2.2 summarize the analysis of the inattentiveness measures for this study. From this it can be seen that there is a significant difference between the two conditions and that the students are more attentive when studying the programmed materials. However, there is no difference among the sessions and no indication of a trend over time under either of the conditions.

Inattentiveness Among Fourth-Grade Students Using an Arithmetic Program

Five fourth-grade classes in three different schools were observed while they were using the multiplication and division program and also while they were doing seat work with material that was not programmed. The classes used this program on a daily basis and devoted about 45 minutes per day to this type of study. Students were permitted to work at their own pace and approximately six weeks were spent on the program. The classes were observed at the end of the first, second, fourth, and sixth weeks. The schedule of observations for this particular study involved observing the classes on two consecutive days at the four times indicated. Each inattentiveness measure is therefore an average for two consecutive days. The classes were also observed on these same days while they were doing some non-programmed

Table 2.1

Mean Inattentiveness Measures Under Programmed and Non-Programmed Study Conditions for Six First-Grade Classes Using the Time-Telling Program

Study Condition	Mean Percent of Students Inattentive				
	First Session	Fifth Session	Ninth Session	Fourteenth Session	All Sessions
Programmed Work	5.42	5.15	4.71	5.27	5.14
Non-Programmed Work	10.90	13.81	10.97	14.08	12.44
Both Conditions	8.16	9.48	7.84	9.68	8.79

Table 2.2

Summary of Analysis of Variance of Inattentiveness Measures for Six First-Grade Classes Using the Time-Telling Program

Source	Sum of Squares	df	Mean Square	F
Prog. vs Non-Prog.	6.4020	1	6.4020	41.26**
Error (Bet. Cl.; pooled)	1.5508	10	.1551	
Among Sessions	.3074	3	.1025	Less than 1
Sessions X Prog.-Non-Prog.	.2563	3	.0858	Less than 1
Error (Sess X Cl.; pooled)	3.4879	30	.1163	
Total	12.0054	47		

**Significant at .01 level

seat work. This design again permitted a study of differences in attentiveness under programmed and non-programmed conditions and a study of changes over time.

Tables 2.3 and 2.4 summarize the analysis of the inattentiveness measures for this study. From this it can be seen that there is a significant difference in inattentiveness between the programmed and non-programmed conditions and that there is less inattentiveness when programmed materials are being used. Also, the difference among sessions is significant. It can be seen that here there is a definite trend, under both conditions, toward an increase in inattentiveness with the passage of time, and that the interaction of sessions and treatments is not significant.

Inattentiveness Among Fourth-Grade Students Using a Spelling Program

Six fourth-grade classes in four different schools were observed while they were studying the spelling program. These classes used this program for an entire semester, devoting about 15 to 20 minutes per day for five days a week to the study of spelling. Three days each week were devoted to the study of the program, with one day used for testing, and the other for teacher-directed "enrichment" activities. Each pupil was required to cover the same number of frames per week but was permitted to progress through the week's work at his own pace. The classes were observed at the end of the first week, at the end of the second week, at the end of the first month, and at the end of the semester. Each class was also observed on these same days during a period when pupils were doing non-programmed seat work. Data were analyzed to determine differences in inattentiveness between the programmed and non-programmed situations and any changes during the course of the semester.

Tables 2.5 and 2.6 summarize the analysis of the inattentiveness measures for the classes that used the spelling program. From this it can be seen that inattentiveness is significantly greater under the condition of non-programmed study. In this situation there was no evident trend in inattentiveness as the semester progressed.

Table 2.3

Mean Inattentiveness Measures Under Programmed and Non-Programmed Study Conditions for Five Fourth-Grade Classes Studying an Arithmetic Program

Study Conditions	Mean Percent of Students Inattentive				
	End of 1st week	End of 2nd week	End of 4th week	End of 6th week	All Sessions
Programmed Work	3.72	5.18	6.90	9.36	6.29
Non-Programmed Work	9.04	9.74	9.58	11.42	9.95
Both Conditions	6.38	7.46	8.24	10.39	8.12

Table 2.4

Summary of Analysis of Variance of Inattentiveness Measures for Five Fourth-Grade Classes Studying an Arithmetic Program

Source	Sum of Squares	df	Mean Square	F
Prog. vs Non-Prog.	1.3359	1	1.3359	11.8**
Error (Bet. Cl.; pooled)	.9054	8	.1132	
Among Sessions	.8631	3	.2877	10.7**
Sessions X Prog.-Non-Prog.	.1771	3	.0590	2.19
Error (Sess X Cl.; pooled)	.6467	24	.0269	
Total	3.9282	39		

**Significant at .01 level

Table 2.5

Mean Inattentiveness Measures Under Programmed and Non-Programmed Conditions for Six Fourth-Grade Classes Using the Spelling Program

Study Conditions	Mean Percent of Students Inattentive				
	End of 1st week	End of 2nd week	End of 1st month	End of Semester	All Sessions
Programmed Work	4.30	3.73	3.10	4.85	4.00
Non-Programmed Work	10.02	9.15	8.50	9.78	9.36
Both Conditions	7.16	6.44	5.80	7.32	6.68

Table 2.6

Summary of Analysis of Variance of Inattentiveness Measures for Six Fourth-Grade Classes Using the Spelling Program

Source	Sum of Squares	df	Mean Square	F
Prog. vs Non-Prog.	3.4381	1	3.4381	26.17**
Error (Bet. Cl.; pooled)	1.3139	10	.1314	
Among Sessions	.1578	3	.0526	Less than 1
Sessions X Prog.-Non-Prog.	.0276	3	.0092	Less than 1
Error (Sess X Cl.; pooled)	2.3408	30	.0780	
Total	7.2782	47		

**Significant at .01 level

Discussion: Studies of Pupil Inattentiveness

It would probably be impossible to design a study to answer the general question of whether programmed materials are more or less effective in holding pupil attention than are other study materials. This question can be answered only in terms of specified programs and specified alternative materials.

However, the three studies reported here do provide some evidence with respect to the question. Here three different programs were involved and three different plans for using the materials were followed.[2] In all three cases the students were significantly more attentive when studying the programs. Also, in two of the three studies there was no trend toward inattentiveness as students became more accustomed to studying from the program. In the one case where students became more inattentive with the passage of time this increase was seen both when they were studying programmed and non-programmed materials.

Taken together, the studies would seem to indicate that some programmed materials can be more effective than non-programmed materials in holding pupil attention and that pupils do not necessarily become less attentive as they continue to use programs over a period of time.

Studies of Pupil Attitude[3]

Further studies of pupil reaction to the use of programmed materials were carried out by obtaining measures of pupil attitude. In these studies measures were obtained both of (1) pupil attitude toward the use of programmed materials and (2) pupil attitude toward the subject being studied. With the first measure it was possible to note pupil changes in attitude over an extended period during which they were studying from programs. With the second measure it was possible to compare attitudes and changes in attitude of pupils studying from programs and pupils studying from non-programmed materials.

---

[2] These plans, described briefly in this chapter, are explained in more detail elsewhere in this report.

[3] The studies reported in this section plus additional studies of pupil attitude are presented in detail in Ferderbar, Joseph E. Changes in Selected Student Attitudes and Personality Measures and Their Relationship to Achievement, Intelligence, and Rate When Using Programmed Instruction, (Unpublished Ed.D dissertation, University of Pittsburgh, 1963).

To obtain these measures of attitude, two separate Likert-type attitude scales of 12 items each were developed. One measured attitude toward the use of programmed materials. The other measured attitude toward the subject. Prior to using these instruments the reliability of each was determined through the use of the split-half procedure and the Spearman-Brown formula. Data from the reliability studies may be summarized as follows:

Scale	N	Reliability Coefficient
Attitude toward Programmed	264	.950
Attitude toward Subject	168	.950

Changes in Attitude Toward the Use of Programmed Materials

A study of what happens to student attitude toward the use of programmed materials as pupils use such materials over the course of a semester was made by measuring attitudes of pupils in four seventh-grade general science classes using a general science program.[4] All pupils involved were asked to respond to the attitude scale at (1) the end of the first week, (2) the end of the second week, (3) the end of the fourth week, and (4) the end of the semester. A summary of these results and an analysis of the significance of the differences found is presented in tables 2.7 and 2.8.

From this summary and analysis it can be seen that with this program used under these conditions there was a definite decrease in favorableness of attitude toward the use of programmed materials. Also, there was

---

[4] These classes worded in general science programmed texts for one semester and studied the following topics: measurement, meteorology, astronomy, sound, light, electricity, and communications. These were topics usually included in the existing seventh-grade curriculum.

Table 2.7

Mean Attitude-Scale (Attitude Toward Use of Programmed Materials) Scores Obtained at Given Sessions for Four Seventh-Grade Classes Using a General Science Program

	Mean Attitude Scale Score for Class				
Class	End of 1st week	End of 2nd week	End of 4th week	End of Semester	All Tests
A (N=31)	45.97	39.94	38.65	30.65	38.80
B (N=32)	48.38	46.16	43.13	30.06	41.93
C (N=32)	29.00	29.50	25.69	24.13	27.08
D (N=31)	43.55	39.36	33.94	27.58	36.11
All Classes	41.73	38.74	35.35	28.11	35.98

Table 2.8

Summary of Analysis of Variance of Attitude-Scale Scores for Four Seventh-Grade Classes Using General Science Program

Source	Sum of Squares	df	Mean Square	F
Sessions	13,670	3	4,556.67	22.26**
Classes	16,504	3	5,501.33	26.88**
Residual	1,844	9	204.67	
Total	32,018	15		

**Significant at .01 level

a significant difference among these four classes in their attitudes toward the use of such materials.

The Correlation of Attitude with Achievement and Intelligence

The measures of attitude toward the use of programmed materials were studied further by determining the correlation of attitude with intelligence and with various measures of achievement. These correlations are presented in Table 2.9. The only thing to note here is that none of these correlations are significant at the .01 level. This indicates that a student's attitude toward the use of programmed materials is not related to his level of intelligence. Also, his attitude does not appear to be a factor in determining how well he learns from a program.

Table 2.9

Correlation of Attitude Toward Use of Programmed Materials (Measured at End of Semester) With Intelligence and Various Measures of Achievement

Test	Correlation with Attitude toward Program (End of Semester)
Otis I. Q.	.025
Coop. Science Test	
Pre-testing	.102
Post-testing	.216
Tests for Units of the Program	
Meteorology	.091
Astronomy	.223
Sound	.202
Light	.152
Electricity	.226
Communication	.159

r of .229 needed for significance at .01 level

Changes in Attitude Toward the Subject Being Studied

Seventh-Grade Science. The relationship of the use of programmed materials to a student's attitude toward the subject he is studying was another aspect of the investigation of attitudes. This relationship was first studied with the students in the four seventh-grade classes studying the general science program. In this case the attitude measures were also obtained for seventh-grade students using non-programmed materials to study general science. All students were given the special 12-statement, Likert-type attitude scale at the beginning of the semester and again at the end of the semester.

The summary of these measures for the four classes studying from the program and the results from the analysis of variance are presented in Tables 2.10 and 2.11. From the summary it can be seen that there was a significant decrease in favorableness of attitude toward the subject during the semester of study. This decrease was present in the case of all four classes involved in the study.

The change in attitude toward the subject was also measured for a group of 63 students studying this same subject, general science, with non-programmed materials. These results are presented in Table 2.12. Here the decrease in favorableness of attitude was also present but was not statistically significant. It should be pointed out that the design of this analysis did not permit a test of the significance of the difference in change in attitude between students studying from a program and those using non-programmed materials.

The measures of attitude toward the subject were investigated further by determining the correlation of attitude at the end of the semester with IQ and with achievement. These correlations are shown in Table 2.13. It can be seen that while none of these correlations are large, most of them are significant at the .01 level.

Table 2.10

Mean Attitude-Scale (Attitude toward Subject) Scores Obtained at Beginning and End of Semester for Four Seventh-Grade Classes Using General Science Program

Class	Mean Attitude-Scale Score for Class		
	Start of Sem.	End of Sem.	Mean
A (N=31)	47.36	36.00	41.68
B (N=32)	49.16	39.69	44.43
C (N=32)	37.81	24.60	31.21
D (N=31)	41.10	38.94	40.02
All Classes (N=126)	43.85	34.76	39.33

Table 2.11

Summary of Analysis of Variance of Attitude-Scale (Attitude toward Subject) Scores for Four Seventh-Grade Classes Using the General Science Program

Source	Sum of Squares	df	Mean Square	F
Sessions	5,158	1	5,158	13.54*
Classes	6,212	3	2,071	5.43
Residual	1,143	3	381	
Total	12,513	7		

*Significant at .05 level

Table 2.12

Mean Attitude-Scale (Attitude toward Subject) Scores Obtained at Beginning and End of Semester for 63 Seventh-Grade Students Studying General Science with Non-Programmed Materials

Mean Attitude Scores		Difference Between Means	t-value for Difference
Start of Sem.	End of Sem.		
48.51	46.25	2.26	1.90*

*Not significant at .05 level with 2-tail test

Table 2.13

Correlation of Attitude Toward the Subject Being Studied (Measured at End of Semester) With Intelligence and Various Measures of Achievement

Test	Correlation with Attitude toward Subject	
	For Prog. Students	For Non-Prog. Students
Otis I. Q.	.121	.323**
Coop. Science Test		
Pre-testing	.174	.349**
Post-testing	.334**	.336**
Tests for Units of the Program		
Meteorology	.202	(Unit tests not given)
Astronomy	.323**	
Sound	.252**	
Light	.224	
Electricity	.346**	
Communication	.317**	

**Significant at .01 level

Programmed Algebra. Using the same scale for measuring attitude toward a subject that was used in the study of attitudes of seventh-grade science students, measures were obtained at the beginning of the school year and at the end of the school year for two classes of ninth graders using an algebra program. This was the algebra program, described more fully in other sections of this report, developed to teach a rather traditional algebra content.

A summary of the attitude measures obtained for these classes and the results from the analysis of variance are presented in Tables 2.14 and 2.15.

Here it can be seen that there was a significant decrease in favorableness of attitude toward the subject from the beginning to the end of the year. It is also interesting to note that the change in attitude was essentially the same for the two classes.

Programmed Modern Mathematics. Change in attitude toward the subject during the course of a school year was also investigated with a class of 36 students using a modern mathematics program. This program has been described in a previous section of this report and is one designed to teach much of the "modern mathematics" content now being recommended for use in a beginning high school level course.

A summary of the attitude measures obtained at the beginning and end of the year for this class and the results of the test of the significance of the change in mean measures are presented in Table 2.16.

With the class and this program, the results were different from what they were in the case where pupils were studying the algebra program. With the modern mathematics program, there was essentially no change in student attitude toward the subject.

Non-Programmed Algebra. To obtain some idea of what happens to pupil attitude toward a subject when the subject is being studied without the use of programmed materials, the same scale used in the above studies was administered at the beginning and the end of the year to 60 students studying with ninth-grade algebra from a rather conventional textbook. The content covered in this course was essentially the same as that covered with the programmed algebra course.

Table 2.14

Mean Attitude-Scale (Attitude toward Subject) Scores Obtained at Beginning and End of Year for Classes Using the Algebra Program

Class	Mean Attitude-Scale Score for Class		
	Beg. of Year	End of Year	Mean
J (N=64)	40.09	28.88	34.39
K (N=67)	40.02	28.64	34.33
Both Classes (N=131)	40.05	28.76	34.40

Table 2.15

Summary of Analysis of Variance of Attitude-Scale (Attitude toward Subject) Scores for Ninth-Grade Classes Using the Algebra Program

Source	Sum of Squares	df	Mean Square	F
Sessions	8,401	1	8,401	70.01**
Classes	42	1	42	.35
Residual	120	1	120	
Total	8,563	3		

**Significant at .01 level

Table 2.16

Mean Attitude-Scale (Attitude toward Subject) Scores Obtained at Beginning and End of Year for 36 Ninth-Grade Students Using the Modern Mathematics Program

Mean Attitude Scores		Difference Between Means	t-value for Difference
Beg. of Year	End of Year		
32.78	37.44	0.16	.08*

*Not significant at .05 level

A summary of the attitude measures for these algebra students studying with non-programmed materials is presented in Table 2.17.

Table 2.17

Mean Attitude-Scale (Attitude toward Subject) Scores Obtained at Beginning and End of Year for 60 Ninth-Grade Students Studying Algebra With Non-Programmed Materials

Mean Attitude Scores		Difference Between Means	t-value for Difference
Beg. of Year	End of Year		
42.48	39.07	3.41	2.26*

*Significant at .05 level

Here again there is a significant decrease in favorableness of attitude during the course of the year. While this decrease is not as large in absolute value as that shown by students using the algebra program, it is a difference which is significant at the .05 level.

Correlation of Attitude Measures with Achievement and Intelligence. The measures of attitude toward the subject for the ninth-grade mathematics students were investigated further by determining the correlation of the measures obtained at the end of the year with selected achievement measures, with intelligence, and with rate of progress through a program. These correlations are summarized in Table 2.18. It will be noted that most of the correlation coefficients are not significantly different from zero and that even those that are significant are quite small.

Table 2.18

Correlations of Attitude Toward the Subject Being Studied (Measured at End of Year) with Intelligence, Rate, and Various Measures of Achievement for Selected Ninth-Grade Students Studying Algebra and Modern Mathematics

Test	Correlation with Attitude Toward Subject		
	Temac (N=131)	SRA (N=36)	Non-Program Students (N=60)
Otis I. Q.	.069	.065	.050
Modern Math	.107	.153	.228
Coop. Algebra Test	.004	.233	.475**
Rate			
9 weeks	-.054	-.009	
18 weeks	.076	-.132	(Other measures not applicable)
27 weeks	-.017	.116	
36 weeks	-.010	.059	
Frame Rate (SRA only)		-.094	
Unit Tests			
First test	.068	.041	
Last test	.272	-.178	

**Significant at .01 level

Discussion: Studies of Pupil Attitude

The studies of pupil attitude presented in this report provide certain descriptive data concerning what happens to the attitudes of students as they study certain programmed materials in regular school situations. The results would seem to offer some suggestions for things that should be of concern to persons or agencies developing programs and for teachers who are using such materials. They also suggest several hypotheses for further study.

In summarizing the data on changes in attitude toward the use of programmed materials it would have to be said that in the situations investigated in this study there was a rather consistent decrease in favorableness of attitude as programs were used over a year's time. Of course, a weakness of the present study was that no comparable measure of change in attitude toward other methods of study was obtained. That is, it might be that if students were questioned at periodic intervals over the course of a school year concerning their liking for any particular method of study that the favorableness of their attitude would steadily decrease.

The investigation of changes in attitude toward the subject being studied revealed no consistent pattern in such changes. That is, the students studying the programmed algebra showed a decrease in favorableness of attitude, those studying programmed modern mathematics showed no change, and those studying algebra with non-programmed materials showed a decrease in favorableness. The interesting thing to note here was that in the two cases where decreases were found, the same subject, algebra, was involved while in the case where interest was maintained, a different and relatively new subject-matter content was being studied. This suggests the very reasonable hypothesis that the content being studied may be a more important determiner of attitude toward a subject than is mode of presentation.

Studies of the Correlation of Achievement and Aptitude[5]

The introduction of programmed materials into the classroom on any broad scale will undoubtedly require a re-examination of the concept of "aptitude" and the relationship of aptitude and achievement. For example, it may be that the typical positive correlation between intelligence and achievement in academic subjects is largely a function of variations in the speed with which pupils can master material. Might not this correlation be expected to be reduced when pupils are permitted to progress at their own rates and where each student masters the material at each step before proceeding to the next, as should be the case when programmed materials are used? In this latter situation all students should attain a high degree of mastery and scores on tests designed to cover a given unit should vary little from low to high intelligence levels. On the other hand, if programmed material is used and each pupil is permitted to progress at his own pace, these variations in amount of material covered could result in an even greater variation of scores on comprehensive measures such as standardized achievement tests. This could, conceivably, result in a higher correlation between aptitude and achievement. It is problems of this type that are the general concern of the studies outlined in this section.

Related Research

The analysis of preliminary data that have been gathered as a part of the present over-all study indicate that, in the case of junior high school students studying general science through the use of a program as compared with students studying general science with non-programmed materials, there is a greater correlation between intelligence and achievement for the latter group than for the former. This seems to be in agreement with the findings of Porter (1959) when spelling was the subject being studied and also with the finding of Meyer (1960) who was studying the results of the use of a vocabulary program.

---

[5] The studies reported in this section plus additional studies involving the correlation of achievement and aptitude are presented in detail in McCormick, James H. Differences in the Relationship Between Achievement and Selected Measures of Aptitude under Programmed and Non-Programmed Instruction (Unpublished Ed.D. dissertation, University of Pittsburgh, 1963).

Correlation of Aptitude and Achievement for Fourth-Grade Students Using Spelling and Multiplication and Division Programs

The correlation between intelligence and aptitude was studied in the case of fourth-grade students using the spelling program and the multiplication and division program. In both cases the correlations obtained with students using a program were compared with those obtained with a similar group using non-programmed materials.

Two hypotheses were of interest here. One was that since using programmed materials should permit all students to master a subject, the correlation of intelligence and achievement should be smaller when programmed materials are used. The second hypothesis was that when pupils study from a program the correlation of intelligence with achievement as measured by a standardized test should be higher than the correlation with a test specifically developed to measure acquisition of the material taught by the program. This hypothesis is derived from the idea that a standardized test will measure more general abilities, some involving transfer of learning, and that intelligence will be a factor in this ability to transfer or generalize.

The studies described here, due to limitations in sampling as well as other factors, cannot be considered as providing any formal test of these hypotheses. They are intended rather to provide objective evidence which should be helpful in planning a more formal testing of hypotheses.

Tables 2.19 and 2.20 summarize the correlations obtained in the two investigations carried out to obtain information on these relationships.

In Table 2.19 the evidence concerning the difference between program and non-program groups in the correlation between intelligence and achievement is inconclusive. As can be seen, with the program test the correlation is higher for the non-program group, with the Iowa Test the correlations are essentially equal, and with the Stanford Test the program group shows the higher correlation. There is no clear evidence here to suggest that the use of programmed materials results in a lower correlation between aptitude and achievement. However, there is evidence to suggest that with pupils studying from a program the correlation of intelligence and achievement will be lower when a special program test is used as a measure of achievement than when a standardized test is used for the latter measure. Also, it seems reasonable to attribute the higher correlation of IQ with the Program test of the non-program group to the fact that the brighter students are more likely to have learned words other than those specifically taught.

In Table 2.20 there is some indication that the correlation between intelligence and achievement is greater for pupils studying from non-programmed materials than it is for pupils using a program. There is also evidence to suggest that with pupils studying from a program the correlation between intelligence and achievement will be greater when a standardized test is used to measure achievement.

The Correlation of Intelligence and Achievement for Seventh-Grade General Science Students

A further investigation of the correlation between intelligence and achievement when pupils study from programmed materials as compared with the situation where pupils study from non-programmed materials was made by studying this relationship for the seventh-grade general science students. Here correlations were determined for the 126 students studying the general science without the use of programmed materials.

In this investigation the Otis test was used to measure intelligence and the Cooperative General Science test was used to measure achievement. Also used as achievement measures for the students using the program were six separate tests developed to measure achievement on each of six units of the program and a special "criterion test" developed to measure over-all achievement on the total program. The inclusion of these program tests permitted a comparison of the correlation of intelligence with such tests designed specifically to measure the objectives of a program and the correlation of intelligence with a more general measure of achievement, the standardized test. The correlations obtained are reported in Table 2.21.

Table 2.19

Correlation of Otis IQ with Selected Measures of Achievement for Fourth-Grade Pupils Using the Spelling Program and for a Comparable Group of Fourth Graders Studying Spelling with Non-Programmed Materials

Group	Correlation of Otis IQ With		
	Program Test	Iowa Test: Spelling Score	Stanford Test: Spelling Score
Program Group (N=169)	.232	.396	.414
Non-Prog. Group (N=38)	.479	.402	.121

Table 2.20

Correlation of Otis IQ with Selected Measures of Achievement for Fourth-Grade Pupils Using the Multiplication and Division Program and for a Comparable Group Studying Multiplication and Division with Non-Programmed Materials

Group	Correlation of Otis IQ With	
	Program Test	Stanford Arith. Av.
Program Group (N=173)	.202	.314
Non-Prog. Group (N=59)	.262	.440

Table 2.21

Correlations of Otis IQ with Various Measures of Achievement for Seventh-Grade Students Studying General Science from Programmed and Non-Programmed Materials

Tests	Correlation with Otis IQ
For students using program (N=126)	
Coop. General Science Test	.532
Program Criterion	.566
Program Unit Tests	
Meteorology	.595
Astronomy	.522
Sound	.470
Light	.415
Electricity	.530
Communication	.576
For students not using program (N=63)	
Coop. General Science Test	.594

A comparison of the correlations between Otis IQ and Coop. test score for the two groups (.532 and .594) indicates that the difference between them is of no practical significance. The same can be said for the difference for the program group between the correlation of Otis IQ with achievement on the Coop. test and with achievement on the criterion test (.532 and .566).

It will be noted that the data obtained for seventh-grade science students concerning the relationships of intelligence and achievement does not show the same differences found in the previously presented data on fourth-grade students studying spelling or arithmetic.

Discussion: The Correlation of Achievement and Aptitude

This section has presented data with respect to the hypotheses (1) that the correlation between IQ and achievement will be lower when students study from programmed materials than when they study from non-programmed materials and (2) that the correlation between IQ and achievement will be greater when a standardized achievement test is used than when the achievement test is one specifically developed for the program of study. In only one of the three studies was the first hypothesis supported, but two of the three studies provided data that supported the second hypothesis. These results, then, provide no conclusive evidence concerning the effect of programmed instruction on the relationship between IQ and achievement. The effect of different methods of study on the relationship of aptitude and achievement is one which should continue to be of concern to educators. However, the results do provide some substantiation for the idea that IQ is not as much a determiner of how well a person can master the specific content in which he is given instruction as it is a determiner of how well he can generalize from what he has been taught to a comprehension of related ideas. This finding, if true, would have important implications for the methods of instruction that should be used with students at different ability levels.

Summary

This chapter has presented data concerning the impact of programmed instruction on a variety of student variables. Three separate studies involving a structured observation of the attentiveness of pupils under programmed and non-programmed instruction provided evidence that pupils were significantly more attentive while studying from programs. Also, there was no general tendency for this attentiveness to decrease significantly over time. However, in separate studies of pupil attitude toward the use of programmed instruction there was a tendency for the pupils' expressed attitude to decrease in favorableness as the programs were used over an extneded period of time. Unfortunately, the studies did not provide for data on comparable changes in attitude when pupils used other methods of study. Three investigations involving changes in attitude toward the subject being studied resulted in data that suggest that the content involved has more influence on attitude than does the method of presentation. Studies of the correlation of IQ and achievement indicate that the use of programmed instruction has little effect on this relationship but that this correlation is slightly higher when achievement is measured by a standardized test rather than a test developed specifically for the unit of study.

References

Ferderbar, J. F. Changes in selected student attitudes and personality measures and their relationship to achievement, intelligence, and rate when using programmed instruction. Unpublished doctoral dissertation, Univer. of Pittsburgh, 1963.

McCormick, J. H. Differences in the relationship between achievement and selected measures of aptitude under programmed and non-programmed instruction. Unpublished doctoral dissertation, Univer. of Pittsburgh, 1963.

Meyer, Susan R. Report on the initial test of a junior high-school vocabulary program. In A. A. Lumsdaine and R. Glaser (Eds.), Teaching machines and programmed learning. Washington, D.C.: National Education Association, 1960. Pp. 229-246.

Porter, D. Some effects of year-long teaching machine instruction. In E. H. Galanter (Ed.), Automatic teaching: the state of the art. New York: Wiley, 1959. Pp. 85-90.

PRECEDING PAGE BLANK- NOT FILMED

CHAPTER 3

PROGRAMMED INSTRUCTION: SOME UNANTICIPATED CONSEQUENCES

Richard O. Carlson[1]

Introduction

Any change is made in anticipation of specified results. In this case of the introduction of programmed instruction, the anticipated results centered on facilitating the learning process in children. The bulk of this report deals with an evaluation of the extent to which programmed instruction fulfills these anticipated results.

Just as change has anticipated consequences, it has unanticipated consequences. The results of the introduction of programmed instruction (or any other innovation) spill over that which was anticipated or intended. It is the purpose of this chapter to report on selected aspects of the spillover, or the unanticipated consequences of the use of programmed instruction.

It is possible to get bogged down on the question of what was intended and what was not. Part of the difficulty is that the acceptors of a new innovation tend to assert that all results were at least party suspected in the beginning. Therefore, they are frequently an unreliable source on the matter of what was intended and what was not. Further, no useful purpose is served here by attempting to draw a fine line between anticipated and unanticipated consequences. The perspective taken in this chapter on the question of what was intended and what was not is arbitrary, and is that all consequences not directly linked to the learning process were unintended.

The reporting of unanticipated consequences involves another potential problem, which is the assessment in terms of their good or bad effects. Making this assessment calls for data well beyond that reported here and, in addition, calls into play an array of personal values that can and do vary greatly. Therefore, no judgments will be attempted here. Nevertheless, this is an extremely relevant question that must be answered by a school district as it evaluates its work.

---

[1] Now at the Center for the advanced study of Educational Administration at the University of Oregon.

Having made it clear that the intention here is neither to prove that certain results were not anticipated nor to assert that these unintended consequences seriously detract or enhance the use of programmed instruction, the task is to simply report some findings that illustrate the range of impact of programmed instruction.

The Method of Evaluation

This impact will be reported as seen through the eyes of the participants who were involved in the programmed instruction experiment. The findings reported here are based on interviews conducted in several stages with building principals, teachers, and central office personnel who had some direct responsibility for programmed instruction. At first the interviews were designed simply to determine the range of changes that these people saw in the alteration of their functions through the introduction of programmed instruction. In most cases, programmed instruction had been in use for a year and a half at the beginning of the interviews so that the informants had some depth of experience to draw upon. As the interviewing progressed, the question changed in order to obtain specific information. The areas of impact of programmed instruction reported here can be divided roughly into three categories: (1) supervision, (2) rates of individual achievement, and (3) the teachers' need to perform.

Classroom Observation and Teacher Supervision

The building principals in Baldwin-Whitehall, like most principals, are responsible for observing and judging the work of the teachers in their building. This responsibility calls for the principal to submit a report to the superintendent on the teaching performance of each teacher, a task for which the principal is by no means unprepared. Schools of education offer or require courses in supervision and classroom observation, credential requirements include such courses, and many books treat the subject. Further, principals tend to see classroom observation as an important area, for they often complain that their time for it is eroded away by other less crucial duties.

In spite of all of this, programmed instruction seems to have undermined the classroom observation routine or maybe even forced it to a halt. Some rather curious things happen when principals observe teachers who are using programmed instruction--curious in the sense that what takes place in no way resembles the practices they use in observing the regular classroom situation.[2] Two forms of classroom observation of programmed instruction seem to take shape. One form simply involves ignoring or abandoning the observation of the teachers using programmed instruction. Approximately one-third of the principals employed this form. Their justification for so doing was that programmed instruction was experimental so they wanted to give the teacher a free hand. After stating their position they were ready to point out the difficulties involved in thinking about how to supervise teachers using programmed instruction. These difficulties are made apparent in the next section. It seemed that those principals who ignored the observation responsibility did not know how to proceed when new instructional methods were used and so they abandoned this responsibility and fell upon the notion of experimental conditions as a rationalization.

---

[2] That the classroom observation practices used in regard to programmed instruction deviated from the "normal" or usual routine was documented only by responses of teachers (both those involved with programmed instruction and those not involved) and principals.

As can be seen, this interpretation may be conjecture, but it is given some support by the other form of classroom observation employed by the remainder of the principals.

The principals that did perform the supervisory function with teachers using programmed instruction tended to throw aside the procedures they had learned, developed, and used in observing regular classroom teaching. Instead, some of the principals asked the students to tell them how the teacher was doing, others were in the habit of looking in only to see if the teacher was sitting at her desk or passing among the pupils (sitting meant that individual help was not being given and moving around the room meant the opposite), and still other principals who observed teachers using programmed instruction made judgments about the teacher's ability by noting her skill in storing the instructional material and keeping the machines and programs in good order. Aside from the question of the adequacy of these forms of classroom observation, it is clear that programmed instruction creates a classroom situation with which principals are not prepared to deal. The principal's training in supervision is built around the standard classroom setting in which the teacher is presenting and discussing material with students, setting their tasks, and directing their group activities.

As far as classroom observation is concerned, programmed instruction is characterized by two features which throw the principal and his supervisor tactics off stride. One has to do with the teaching goals. A general strategy of supervision is to require teachers to submit their daily lesson plans to the principal for evaluation on a number of criteria such as adequacy of goals, and variety of activities. The plans also give the principal a means of judging the activities in any class he might observe for he has the instructional objectives available. As can be seen this stratagem is rendered irrelevant with programmed instruction because there are no daily lesson plans to be submitted. In fact the program sets the overall goals, and many of the daily activities were controlled by the pace of the class or individual students and the design of the study which the teacher was required to follow as outlined in her manual.

The second characteristic feature of programmed instruction which undermines the standard practice of classroom observation is that it breaks into a mere collection of individuals that which is a group in the regular classroom setting. In programmed instruction the teacher deals with the students on an individual basis whereas in the regular classroom the teacher deals mainly with the students as a group. Current supervisory tactics are geared to the group setting. An examination of any observation form or textbook will make this clear. So, to the extent that the group situation is absent, the principal finds that he cannot observe many of the activities that he has considered to be central to a good learning situation.

Without daily lesson plans and without seeing the teacher work with a group, the principal seems to be thrown off stride and either ignores the classroom observation function or resorts to asking students to judge the teachers or merely notes her clerical-like abilities. It seems clear that new supervisory procedures must be developed.

As a sidelight on the matter of supervision, it seems that programmed instruction will be an element that will interfere with the opportunities that principals take to visit classrooms. Most principals reported that they themselves went into the classroom to substitute for an absent teacher if she was involved in programmed instruction. This was not the routine way of providing for substitutes, and principals did not serve as substitutes for teachers not involved in programmed instruction. In part this suggests that principals see programmed instruction as the most important activity in the school in the sense that it cannot be entrusted to anyone less qualified than the principal. Also, it suggests that school districts will need to develop specialists who can work with programmed instructional classes among the substitutes.

Rates of Individual Achievement

An important anticipated consequence of programmed instruction is that all pupils will be able to learn at their own rates. However, two additional notations must be made. On the one hand there are forces operating to minimize the differences in individual rates of achievement, and on the other there are school-wide consequences when learning proceeds at vastly different rates.

In a dramatic way programmed instruction forces a school to stand face to face with the fact that students learn at widely varying rates. It is true that some of the most shopworn cliches such as "we teach children, not subjects" and "start the learning experience where the child is" reflect a concern for individual differences and suggest that educators are most anxious to tailor learning needs and speeds to individuals. When faced with programmed instruction, which permits students to work at their own rates, however, the hollowness of the cliches is exposed, and a host of practices emerge in an effort to keep students working at similar rates.

Some of these practices could be classified by the cynic as sabotage. The clearest example of this is that as a program progressed, the levels of individual achievement and the range in number of frames completed varied widely, and, as a means of "correcting" this, several teachers were discovered either consciously or unconsciously to be pacing students. That is, teachers were actually restricting the output of the students proceeding at the fastest rates. The logic of restricting output of fast students is tidy and makes good sense from at least one viewpoint. Explaining the same troublesome point to five students who hav encountered it at the same time is less time consuming than explaining the same troublesome point to the same five students as they encounter it at different points in time. For the teacher who complains that there is never enough time this appears to be most efficient. In fact, insisting that all students move at the same rate, which is attempted in many classrooms, can be supported by the same logic.

But pacing the work of fast students by teachers is not the only way in which the schools have attempted to "correct" the move toward total individualized instruction brought on by programmed instruction. A practice sanctioned by the administration sets to work on the other end of the continuum, the slow student. In the general science program it was deemed wise, in the face of ever-growing differences in accomplishment, to allow some students to work on their programs at home. The original decision of restricting the use of programs to the regular class period was reversed only for slow students; average and fast students continued to have access to the programs only during class time. It could be argued that this procedure of allowing only slow students access to programs outside of class time is a clear example of the school's attempt to individualize instruction. However, average and fast students were not allowed extra-class access to the programs and the procedure had the net effect of minimizing the range of student progress. This is one example of several attempts to "correct" almost total individualized instruction which seems possible with programmed instruction.

In addition to restricting the rate of progress of fast students and allowing slow students more time with the programs, another practice emerged which can also be seen as a "correctional" factor. "Enrichment materials" seem to be used in an effort to keep the range of progress through a program at a minimum. Such material developed around some of the programs was used most extensively with the students making rapid progress through a program; those making slow progress encountered it less often. In a sense, slow students had the task of working through the program, fast students had this responsibility plus that of some level of mastery of the enrichment materials. This mechanism greatly increases the possibility that the level of achievement on the program by the slow student will be equal to the level of achievement on the program by the fast learner in the same number of class days. This is not to say that the use of enrichment material is undesirable, but simply that the use of enrichment material as here indicated tends to bring about a condition of minimum spread of rates of progress through the program.

The point being made here can be summarized as follows. Programmed instruction makes it possible for each student to proceed at his own rate.

In this sense, it allows for individual differences. Programmed instruction, possibly more than any other innovation, presents the teacher and the school with the opportunity of achieving individualized instruction. But when faced with this opportunity, in this single case, mechanisms emerged--such as restricting the pace of fast learners and making time to work on programs more available to slower students--which tended to minimize the spread of progress through the program by students of varying ability. Therefore, it can be said that one of the consequences of the acceptance of programmed instruction is that it brings the educator face to face with a situation in which individual differences can be met; but the educators in the case reported here instituted practices which reduced the range of differences in achievement that might otherwise have been evident. All of this simply suggests that schools as they are now structured are either unable or unwilling to accept something near total individualized instruction.

The second consequence of a program where students can move at widely varying rates is simple and clear: scheduling problems become more complex. This can be seen when the question is raised as to what the students do who either fail to complete the program during the allotted time or finish it far earlier. Given the problem of schools (as they are now constituted) in dealing with such occurence, it seems reasonable to assume that this is one of the possible pressures which lead to the institution of the above-mentioned "corrective" actions.

A variety of solutions were developed to "take care of" students who either finished too early or did not finish. Most of them were focused on the latter case where the student could not complete the work required. In part, this occurred because of the assumption that the able student can more easily be occupied in the learning process. One procedure was to permit students to take two semesters to complete a program designed for one semester. Another practice was to enroll the student in a summer session to enable him to complete a program he did not finish in the allotted time. As can be seen, the procedure of letting slow students take the programs home was in part an attempt to avoid the necessity of either of these alternatives. One of the very interesting solutions consisted of holding students who had not finished some weeks later, and then inserting them in an ongoing speech or music class which they were scheduled to begin in the winter semester. This seems interesting to the extent that it suggests a value priority held by educators for the various subjects that schools offer. In no way did this solution increase the popularity of programmed instruction among speech and music teachers.

The Teacher's Need to Perform

The experience of the Baldwin-Whitehall school district with programmed instruction seems to indicate that teachers have a somewhat compelling need to perform. "Perform" here means (a) capture and hold the attention of a number of students, and (b) serve continuously as the mediator between the student and the information. This is what teachers seem to define as teaching. All other acts seem to be assessed as supporting acts, not teaching. Programmed instruction does not give the teachers as much opportunity to perform as they apparently desire; it does not give them sufficient opportunity to teach. In their eyes, because teaching means performing, using programmed instruction is not teaching.

All of this can be inferred from conversations with teachers and seen in what happened with some of the programs. The mathematics program serves as a good example. At the beginning of the experiment with programmed instruction, students of mathematics met in the usual number with one teacher and proceeded to move through the program individually at their own rates. The second year's use of the mathematics program brought the following innovation: two groups of students were scheduled to meet simultaneously with two instructors and, under this arrangement, students worked on the program part of the time and met with a group of students who had reached approximately the same place in the program and an instructor using regular teaching procedures for another part of the time. Other programs were stretched to last longer than the allotted time in order to enable teachers to spend part of the time performing, serving as "director

of learning"[3], rather than the program serving in this capacity. The justification made by the teachers for these innovations was that the students wanted more interaction with their instructors.

It can be seen that the program replaces the teacher in the role of "director of learning," and the net effect of the innovations or changes introduced by the teachers and administrators was to modify programmed instruction in such a way that it took on more of the characteristics of regular classroom instruction, and permitted them to recapture some aspects of the role they wished to fill. The inventions created by the teachers imply that if the logic of programmed instruction is to have its way in schools, a new definition of what teaching is must be conveyed.

---

[3] This term is taken from Teaching Competence, a report by the Commission on Teacher Education, San Francisco: California Teachers Association, 1957, where it is used to signify the main function of teachers.

Summary

Programmed instruction has an impact on a school system beyond its impact on student achievement. In this chapter three areas illustrating such impact have been discussed. Through interviews with teachers and administrators involved with programmed instruction, it was found that programmed instruction introduces elements into the teaching situation which render the standard methods of classroom observation of teachers by principals inadequate. Programmed instruction also permits students to move through their learning tasks at widely varying speeds. Faced with this situation which is defined as desirable by educators, practices emerged which successfully reduced the variability in student progress. Programmed instruction also replaces the teacher as "director of learning" and, having the apparent need to perform, teachers introduced their own innovations which enabled them to recapture some of the role of "director of learning" which was lost to the program.

PRECEDING PAGE BLANK-NOT FILMED

PART TWO:

PILOT STUDIES OF THE INDIVIDUALIZED CLASSROOM

PRECEDING PAGE BLANK-NOT FILMED

CHAPTER 4

THE USE OF PROGRAMMED MATERIALS TO INDIVIDUALIZE AND EXTEND INSTRUCTION

James R. Hawker[1] and Robert Glaser

Introduction

Two main findings of the studies described in Part One have been that (1) students proceed through the material at widely divergent paces--with many students able to proceed at a pace more rapid than that permitted by conventional classroom practices; and (2) a significant number of students know the materials to be taught before receiving instruction even though it is usually taught at a particular grade level. These results suggest that an important need in the improvement of classroom instruction is the development of procedures which more precisely determine what a student knows when he comes into the instructional situation, and a procedure for adapting the instructional sequence to his particular requirements. Such techniques, of course, involve the development of procedures for the individualization of instruction, and represent an area in which programmed instruction might be used. Consequently, the research program for the 1963-64 academic year was oriented toward carrying out a series of pilot studies designed to increase the individualization of instruction, and, at the same time, to provide opportunities for extended learning. In an attempt to adapt the pace of instruction to the individual student, programmed instructional materials were utilized at various grade levels within a classroom structure that was revised in order to permit a flexible teaching situation. More specifically, a series of five studies were carried out in the third, fourth, and fifth grades to evaluate individualization procedures in the teaching of spelling and various arithmetic topics. In addition, a sixth study was carried out at the fifth grade level in which programmed materials were used as a tool for individualized review of fourth-grade material before entering into new fifth-grade work.

---

[1] Now at the Austin State School and The University of Texas, Austin, Texas.

Techniques for Providing Individualization

The general procedure followed in these studies was to provide carefully selected programmed materials in specific subject matters, to structure the classroom situation so that the student could work through the materials at his own pace, and to permit the student to by-pass those portions of the programmed material which pretest performance measures indicated he already knew. This was accomplished by dividing the material into small teaching units and providing detailed pre- and posttests to assess the student's mastery of the material within that unit. Then, before beginning work on one of the units in the program, the student would take the unit pretest and if his score were above a predetermined criterion, he was allowed to skip that unit of material entirely and proceed to the next unit. If his score were below the criterion, he was required to work through part or all of the unit, and, upon completion of the work, to again take the unit test. The criteria used for determining progress from unit to unit were the following:

1. 90% and above--considered mastery level. The student will proceed to the next unit test.
2. 80-90%--some additional work suggested within the unit. The student may qualify to go on to the next unit of work as his posttest score indicates.
3. 50-79%--extensive work to be done within the program. The student should do all of the frames. Remedial work to be emphasized.
4. Below 50%--lower grade work to be instituted. Tutorial and remedial work to be emphasized.

Thus, the student was allowed to progress through the material at his own pace and was not required to spend time on the material which he already knew. By allowing the student to by-pass that portion of the material in which he indicated mastery, he was able to go on to more advanced topics in which he did need instruction. In this way, it was possible to individualize the instruction to some extent for each student, and to provide an opportunity for extended learning in a particular topic. As the studies proceeded, it became apparent that many of the students were completing the programmed materials ahead of schedule and that there was a need for advanced work in particular topics. As a result, it was necessary for the staff to develop supplementary self-teaching materials so that the individualization procedures could be continued. Using conventional textbook and workbook materials, short teaching units were developed on particular topics, appropriate portions of the text or workbook were identified, and the material was set up so that the student could work on it alone or with minimal help from the teacher. Thus, while these materials did not constitute "programmed" instruction in the strict sense of the word, they did permit the students to continue with a program of individualized instruction.

Design and Evaluation of Studies

The remaining portions of this chapter will discuss each of the studies in detail; however, all of the studies were quite similar in design and in experimental procedure. Groups of students received the individualized programmed instruction as part of their regular curriculum. Unless otherwise dictated by the particular experimental design, each group of students receiving programmed instruction was matched with a similar group of students receiving conventional instruction and who served as a Control Group. The matching of groups was done on variables such as IQ and academic achievement levels, using data collected by the school system at the beginning of the year. In addition, a battery of tests was selected for each grade level and these were given both as pre- and posttests to assess and compare the achievement of the Experimental and Control Groups. Finally, detailed records were kept on the Experimental students' progress through the program, e.g., amount of time worked in each unit, amount of material covered, etc.

Individualized Instruction in Spelling

To follow up the studies described in Chapter 1, two studies were carried out in spelling instruction. In the first study, interest was directed toward moving the spelling program, considered to be fourth-grade work, to the third grade level. Since the earlier study had indicated that many fourth-graders knew the material before beginning the program, it was felt that perhaps the program could profitably be used by selected third-grade students. In the second study, the program was again given to fourth-grade students to see if the individualization techniques would be more effective than the paced technique used previously.

Third-Grade Spelling

Two third-grade classes (total N=66) served as the Experimental Group and received individualized spelling instruction using a programmed textbook, while two additional classes (total N=39) served as a Control Group and received traditional instruction in spelling. An attempt was made to match the two groups as closely as possible on IQ and achievement level; however, the classes came from two different schools and it was not possible to match them in the present study. Table 4.1 presents summary data for the two groups on several standard tests given at the beginning of the study. The groups did not differ significantly in IQ (t=1.63, df=103, p>.05), but on the achievement tests, the Control Group was, in all cases, superior to the Experimental Group. Thus, students in the Experimental Group were somewhat behind those in the Control Group in achievement level at the beginning of the study.

Materials. Students in the Experimental Group were given a programmed textbook to be used for their spelling instruction. This program contains approximately 3,000 frames divided into 12 teaching units. Within the program, the student is required to spell a total of 447 complete words, plus significant parts of hundreds of other words. In addition, the program provides instruction in phonics, syllabication, pluralization, suffixes, prefixes, the use of words in sentences, and so on. As the study progressed, it was found that many students were moving quite rapidly through the program, and additional material was needed to continue the individualization techniques. Center staff members and the classroom teachers worked together and developed supplementary materials using the Multi-Level Speller. This book contains approximately 3700 words and is intended for instructional use with grades 3 through 12. There are 165 lists grouped into eleven spelling levels of increasing difficulty, but the levels do not correspond

Table 4.1

Summary Data on Standard Intelligence and Achievement Tests Given at the Beginning of the Year to Students in the Third Grade Spelling Study

	Experimental Group (N=66)		Control Group (N=39)		Group Comparison
Test	$\overline{X}$	SD	$\overline{X}$	SD	t
Otis Quick Scoring Mental Ability Test (IQ)	117.52	12.72	113.90	9.59	1.63
Metropolitan Achievement Tests: Battery Median Grade Equivalent	3.72	0.70	4.11	0.59	-3.02*
Metropolitan Achievement Tests: Average Reading Grade Equivalent	3.71	0.78	4.15	0.57	-3.31*
Metropolitan Achievement Tests: Spelling Grade Equivalent	3.90	0.86	4.41	0.68	-3.37*
Metropolitan Achievement Tests: Spelling Test	25.23	4.48	27.62	3.71	-2.80*

$^*p < .01$

with particular grade levels. In addition, each spelling level has a placement test, consisting of a sample of 20 words from that level, which can be used to determine the point at which the student should begin work. Since primary interest in the present study was focused on determining the extent to which these students could pursue advanced materials, and not simply to see how many additional words they could learn, it was decided that only the placement lists would be used in developing the supplementary materials. Thus, units were worked out on the eleven placement lists making a total of 220 additional words available for study.

Students in the Control Group received conventional classroom instruction in spelling using regular third-grade spelling textbooks and class exercises conducted by the teacher.

Procedure. At the beginning of the school year, a battery of diagnostic and achievement tests in spelling and language was given to both groups of students to determine spelling achievement. On the basis of these pretests and teacher evaluations, those students in the Experimental Group who appeared ready for the fourth-grade spelling program were permitted to begin work immediately with the programmed materials and were sent to another room which was supervised by a teacher and a teacher-aide during the regular spelling period. Those students whose pretest data indicated that they needed some remedial work remained in the regular spelling class until they had overcome their deficiencies and then went into the programmed material. Thus, all students in the Experimental Group did not begin work in the program at the same time, but began when their performance level indicated they were ready to profit from individualized instruction. As the students completed the programmed materials, they returned to their regular class and began work in the supplementary materials.

Students in the Control Group received traditional spelling instruction in the regular classroom, and the teachers were told to follow their usual teaching practices using any materials which they felt would be beneficial.

Results and Discussion. All students were again given the above battery of tests at the end of the year, and summary data on the pre- and posttest scores are presented in Table 4.2. These data indicate that, except for the Spelling Program Test, the Control Group scored significantly higher than the Experimental Group on all of the tests at the beginning of the study. However, the posttest scores indicate that the two groups did not differ significantly at the end of the year. In terms of gain scores (i.e., the increase from pre- to posttesting), this means that the Experimental Group showed much higher gains on all of the tests (see Table 4.2), and, in fact, all of these gain scores except those on the program test were statistically significant. Thus, the use of programmed materials enabled the students to overcome their initial deficiencies and to raise their general performance level considerably.

In addition to the tests used for pre- and posttesting, another test was developed to provide a measure of a wide range of spelling achievement for all students in the spelling studies. The test consisted of 120 words selected from the New Iowa Spelling Scale and represented a sampling of words from each grade level from grade three through eight. Twenty words were selected from each grade level and were words which had been correctly spelled by approximately 50% of the students in the standardization sample used in developing the scale. Thus the list at each grade level consisted of words of medium difficulty. The groups were compared on each list and on the total number of correct items and were approximately equal on all lists except the Grade 8 list where the Experimental Group scored higher than the Control Group (t=1.99, df=103, $p<.05$). The Experimental Group also had a higher mean number of items correct on the total list, but the differences were not large enough to be statistically significant (see Table 4.2).

In addition to the test data, records were kept of the amount of time required by each student in the Experimental Group to complete the program and the distribution of these data is presented in Figure 4.1. As the figure shows, the fastest students were able to complete the entire program within 31 to 40 days, while the slowest students required 151 to 160 days for completion ($\bar{X}$=78.44, SD=21.72). Thus, it can be seen that the individualization procedure greatly accentuated the individual differences in learning rate among the students.

It is also of interest to compare the amount of material covered by the Experimental and Control groups. The Control Group worked for the entire year (approximately 180 school days) in their regular spelling textbook which introduced 357 new words (compared with 447 in the programmed text), as well as the other material on phonics, syllabication, etc. Thus, if one looks at the amount of material covered and the amount of time spent, it can be seen that the use of programmed materials permitted students to cover approximately 25% more material in an average time , less than half of that required by students receiving conventional instruction. Also, students in the Experimental Group were given advanced work in the Multi-Level Speller after they finished the program. This material, it will be recalled, consisted of eleven 20-word lists of increasing difficulty. Figure 4.2 gives the percentage of students successfully completing the various lists along with cumulative number of words mastered. (In the Botel book, A through K roughly corresponds to grade levels 3 through 12.) As the figure indicates, many students were able to complete a number of additional lists so that the total number of words mastered was considerably above that of the Control Group.

Thus, these data seem to indicate that the individualized use of programmed materials for spelling instruction was quite effective with the present group of third-grade students. The students showed higher gains than the Control Group on all of the achievement tests, and covered considerably more material in the allotted time. In addition, the extent to which the Experimental Group students "spread" themselves out suggests that the average student can cover a greater amount of material than is possible in the conventional classroom, and that even the slower students are able to cover a prescribed amount of material in less time than normally required in the intact classroom.

Fourth-Grade Spelling

Fifty students from two fourth grade classes constituted the Experimental Group, and 50 students from two other classes served as a Control Group. (Unfortunately, however, a large number of students in the Control Group were absent for extended periods of time due to an epidemic of measles and mumps in the school so that the records for these students were not complete for many of the tests which were given. Twenty students fell into this group so that the final size of the Control Group was 30 subjects, and only these subjects were used in making comparisons between the two groups.) The two groups were matched on IQ and general achievement

Table 4.2

Summary Data on Pre- and Posttests for Students in the Third-Grade Spelling Study

Test		Experimental Group (N=66)		Control Group (N=39)		Group Comparison
		$\bar{X}$	SD	$\bar{X}$	SD	t
Spelling for Word Mastery	Pre	13.98	10.90	20.77	11.42	-2.96**
	Post	33.92	6.03	35.23	66.66	-0.99
(40 items)	Gain	19.94	7.77	14.46	8.52	3.25**
Metropolitan Achievement Spelling Test	Pre	13.94	11.66	20.62	10.60	-2.97**
	Post	31.41	8.26	32.28	7.24	-0.56
(40 items)	Gain	17.47	8.09	11.66	5.78	4.15**
Iowa Basic Skills: Spelling	Pre	14.98	6.85	17.82	6.07	-2.18*
	Post	24.05	4.67	24.44	4.78	-0.40
(30 items)	Gain	9.07	5.82	6.62	3.77	2.58*
Spelling Program List A	Pre	10.45	5.52	12.79	5.61	-2.05
	Post	23.06	5.31	19.77	5.29	3.04**
(30 items)	Gain	12.61	3.64	6.98	3.03	8.46***
Spelling Program List B	Pre	10.39	6.72	14.51	6.15	-3.17**
	Post	22.45	5.64	21.79	5.02	0.61
(30 items)	Gain	12.06	3.58	7.28	3.67	6.39***
Spelling Program Test	Pre	31.21	7.37	28.87	7.38	1.55
	Post	42.77	8.22	40.44	9.63	1.25
(69 items)	Gain	11.56	6.71	11.57	7.88	0.14
Iowa Spelling Scale List (120 items)	Total	60.64	19.92	58.85	19.18	0.45

*$p < .05$ (two-tailed test)
**$p < .01$ (two-tailed test)
***$p < .001$ (two-tailed test)

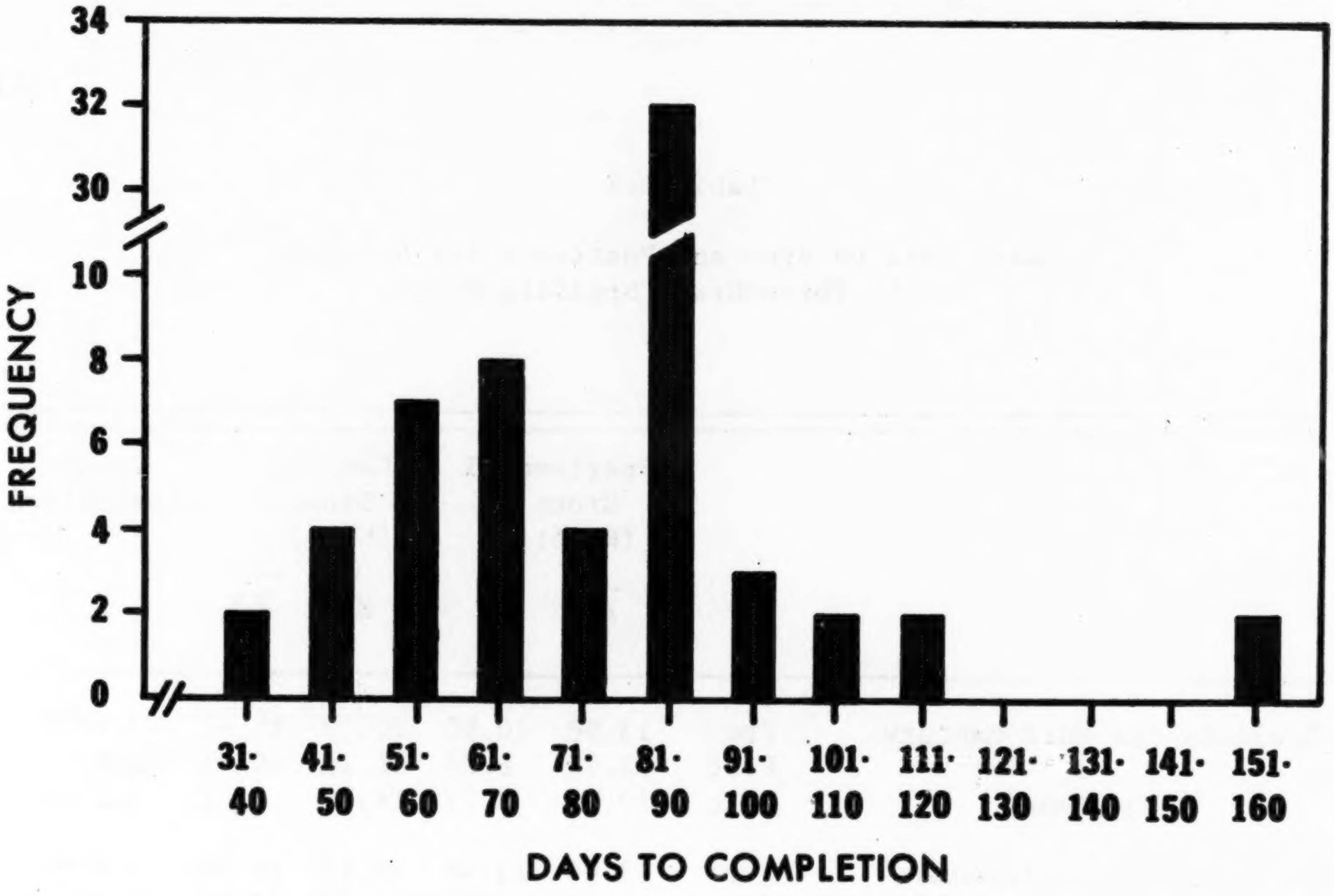

Figure 4.1 Number of days to complete spelling program by third-grade students (N=66).

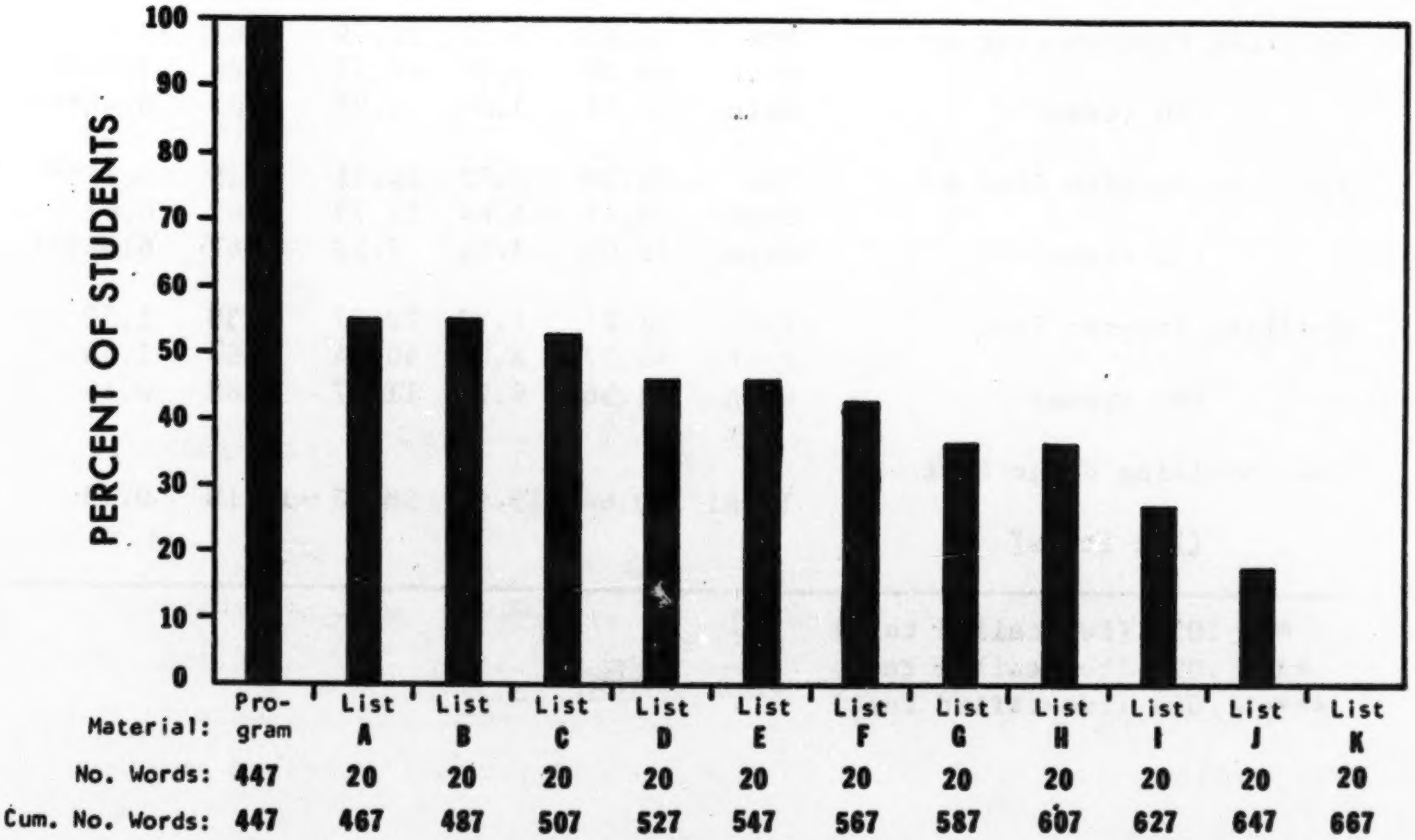

Figure 4.2 Percent of third-grade students completing spelling program and advanced material in spelling (N=66).

tests and, even with the reduction in the size of the Control Group, the groups did not differ significantly on any of the measures (see Table 4.3). Thus the groups began the study with equal achievement in spelling and language subjects.

Materials. The same materials outlined in the third grade study were used in the present study. Students in the Experimental Group received the Spelling Program, and when completing it, moved into supplementary material developed from the Multi-Level Speller. Students in the Control Group used the regular fourth grade spelling textbook and any other materials which the teachers wished to use.

Procedure. The general procedure was essentially the same as that used with the third grade. Both groups were given the battery of diagnostic and achievement tests at the beginning of the year and then went into spelling instruction. Students in the Experimental Group began work immediately in the program and, utilizing the individualization technique described previously, worked through the program on an individual basis and at their own pace. As they finished the program, they began working in the Multi-Level Speller. Unlike the procedure in the third grade, however, the students in the fourth grade Experimental Group were required to work through all of the lists in the Multi-Level Speller. That is, instead of working only on the placement tests at each of the 11 levels, the student would work through all of the lists at each level. Before beginning work on a particular list, the student was tested on it and if his score was sufficiently high (i.e., above the 90% correct criterion), he was considered as having mastered that list and went on to the next list. If his score was not above criterion, he was given exercises involving the words he had missed, and then would take the test again. Thus, the same procedures were used for the supplementary work as for the program work.

The Control Group again received traditional spelling instruction in its regular classroom using textbooks and other conventional materials. At the end of the year, both groups were given the various posttests and the spelling grade range test which was developed and described in the previous study.

Table 4.3

Summary Data on Standard Intelligence and Achievement Tests Given at the Beginning of the Year to Students in the Fourth Grade Spelling Study

Test	Experimental Group (N=66)		Control Group (N=39)		Group Comparison
	$\bar{X}$	SD	$\bar{X}$	SD	t
Otis Quick Scoring Mental Ability Test	111.36	13.08	114.80	14.40	-1.05
Metropolitan Achievement Tests: Battery Median Grade Equivalent	4.89	0.90	4.70	0.82	0.96
Metropolitan Achievement Tests: Average Reading Grade Equivalent	4.81	1.18	4.47	0.89	1.41
Metropolitan Achievement Tests: Spelling Grade Equivalent	5.21	1.05	5.33	1.12	-0.47
Metropolitan Achievement Tests: Spelling Test	30.52	7.49	31.37	7.18	-0.50

Results and Discussion. The pre- and posttest data for each group are presented in Table 4.4, and indicate that the two groups did not differ significantly on any of the pretests except the Spelling Program Test on which the Experimental Group had a higher mean score (t=2.81, df=78, $p<.01$). Similarly, the groups were essentially equal on the posttests with the exception again of the Spelling Program Test, but this time, the order of the means was reversed--that is, the Control Group had a higher mean score than the Experimental Group (t=-2.71, df=78, $p<.01$). The Experimental Group did score higher than the Control Group on the spelling grade range test (means of 71.16 and 66.27, respectively), but the difference was not sufficiently large to indicate a reliable difference in the performance of the two groups (t=0.93, df=78, $p>.05$).

Table 4.4

Summary Data on Pre- and Posttests for Students in the Fourth Grade Spelling Study

Test		Experimental Group (N=50)		Control Group (N=30)		Group Comparison
		$\overline{X}$	SD	$\overline{X}$	SD	t
Spelling for Word Mastery	Pre	35.64	10.62	32.17	10.11	1.44
	Post	43.88	6.96	46.20	5.07	-1.69
(40 items)	Gain	8.24	5.74	14.03	8.07	3.41**
Stanford Achievement Test	Pre	38.18	8.88	37.13	7.30	0.56
	Post	44.20	5.37	44.63	5.32	-0.35
(40 items)	Gain	6.02	4.84	7.50	4.09	1.51
Iowa Basic Skills: Spelling	Pre	22.10	5.64	20.40	4.69	1.43
	Post	26.56	3.51	26.43	3.14	0.16
(30 items)	Gain	4.46	3.69	6.03	4.54	2.11*
Spelling Program List A	Pre	18.42	4.76	17.20	4.55	1.12
	Post	23.78	5.17	24.33	4.00	-0.53
(30 items)	Gain	5.36	2.52	7.13	2.86	2.92**
Spelling Program List B	Pre	20.38	6.12	18.30	6.29	1.43
	Post	24.66	4.54	25.70	4.67	-0.96
(30 items)	Gain	4.28	3.36	7.40	3.57	3.68***
Spelling Program Test	Pre	35.56	10.10	28.73	10.54	2.81**
	Post	44.08	9.17	50.50	10.63	-2.71**
(69 items)	Gain	8.52	6.63	21.77	6.99	7.99***
Iowa Spelling Scale List	Total	71.16	24.59	66.27	21.02	0.93
(120 items)						

*$p < .05$ (two-tailed test)
**$p < .01$ (two-tailed test)
***$p < .001$ (two-tailed test)

The failure of the two groups to reflect any differences in posttest performance is somewhat puzzling, and the fact that the Control Group scored higher than the Experimental Group on the Program Test is especially so. In the previous study, reported in Chapter One, with fourth-grade students, it was found that while the Experimental and Control groups did not differ reliably on standardized tests such as the Iowa Spelling Scale and the Stanford Achievement Test, the Experimental Group did score significantly higher than the Control Group on the Program posttest. The reason for this discrepancy in results is not clear at the present time.

Although the test results do not reflect any differences in end-of-year achievement level, it is interesting to examine the differences in rate of learning among subjects in the Experimental Group and the Control Group. Figure 4.3 presents the distribution of students in the Experimental Group completing the program in different numbers of days, and the wide variability in the amount of time required to complete the program can again be noted. Five of the students finished the program within 21 to 30 days while the slowest two students required 121 to 130 days with the mean number of days required by all students being 66.02 days (SD=31.81). As in the previous study, the Control Group worked the entire school year of 180 days during which time they received 470 new words in their textbook. Thus, the two groups covered approximately the same amount of material, but the Experimental Group completed it in an average time almost a third of that worked by the Control Group. Moreover, students in the Experimental Group completed a large number of lists in the supplementary material, as is shown in Figure 4.4. This figure shows the number of words in each spelling level as well as the total number of words mastered by various percentages of students. For example, almost 50% of the students covered an additional 1,487 new words—approximately three times the number covered by the Control Group. Thus, although the achievement test scores did not reflect significant differences in the performance of the two groups, these data indicate an impressive difference in the amount of material covered by the two groups.

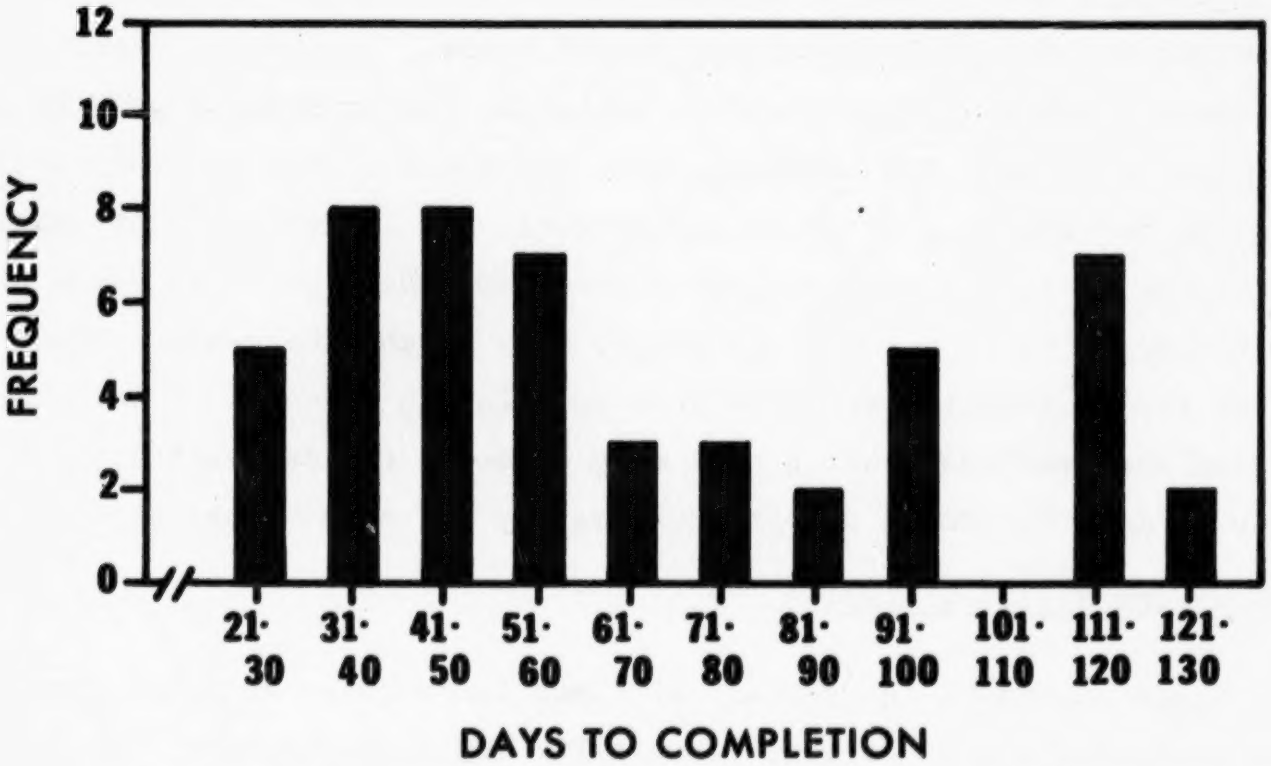

Figure 4.3 Number of days to complete spelling program by fourth-grade students (N=50).

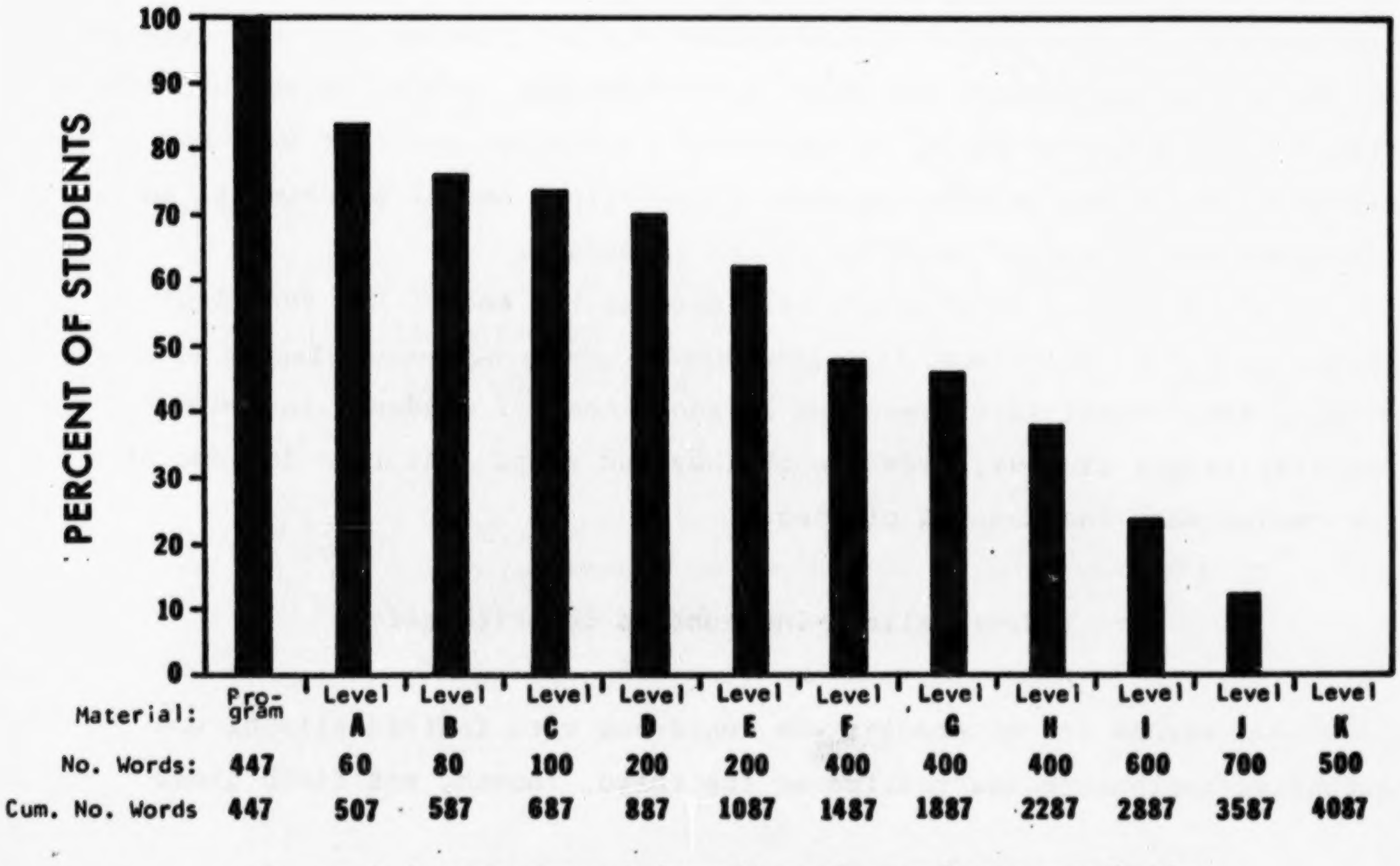

Figure 4.4 Percent of fourth-grade students completing spelling program and advanced material in spelling (N=50).

Finally, it is interesting to compare the rate of learning for the two Experimental Groups in third and fourth grade. Figure 4.5 presents the cumulative percentage of students finishing the programmed work in different numbers of days and indicates that there was a good deal of overlap between the two grades. As would be expected, the fourth grade students completed the program in significantly less time than the third grade students ($\underline{t}=2.37$, $\underline{df}=114$, $\underline{p} < .05$); although, many of the third-grade students completed the program in less time than some fourth graders. These results underscore the previous finding that many students can successfully pursue a course of instruction in a topic customarily set at a higher grade level.

Summary of the Spelling Studies

Taken together, the results of these two studies indicate that programmed materials can be effective tools for individualization. In both studies, students using the programs completed comparable amounts of material in less time than students receiving conventional instruction, and consequently were able to go into advanced work not normally available to them at their particular grade level. Moreover, the extent to which the Experimental Group students spread themselves out implies that the average and above-average student can cover a considerably greater amount of material than is possible in the conventional classroom, and that even the slower students may be able to cover a prescribed amount of material in less time than normally required in the classroom.

Finally, the achievement test data at the end of the year indicated, with few exceptions, that students in the programmed classes reached a final achievement level equal to or above that of students in conventionally taught classes, even though they had spent less time in some of the topics than the control classes.

Individualized Instruction in Arithmetic

A second set of studies was concerned with individualizing and extending arithmetic instruction at the third, fourth, and fifth grade

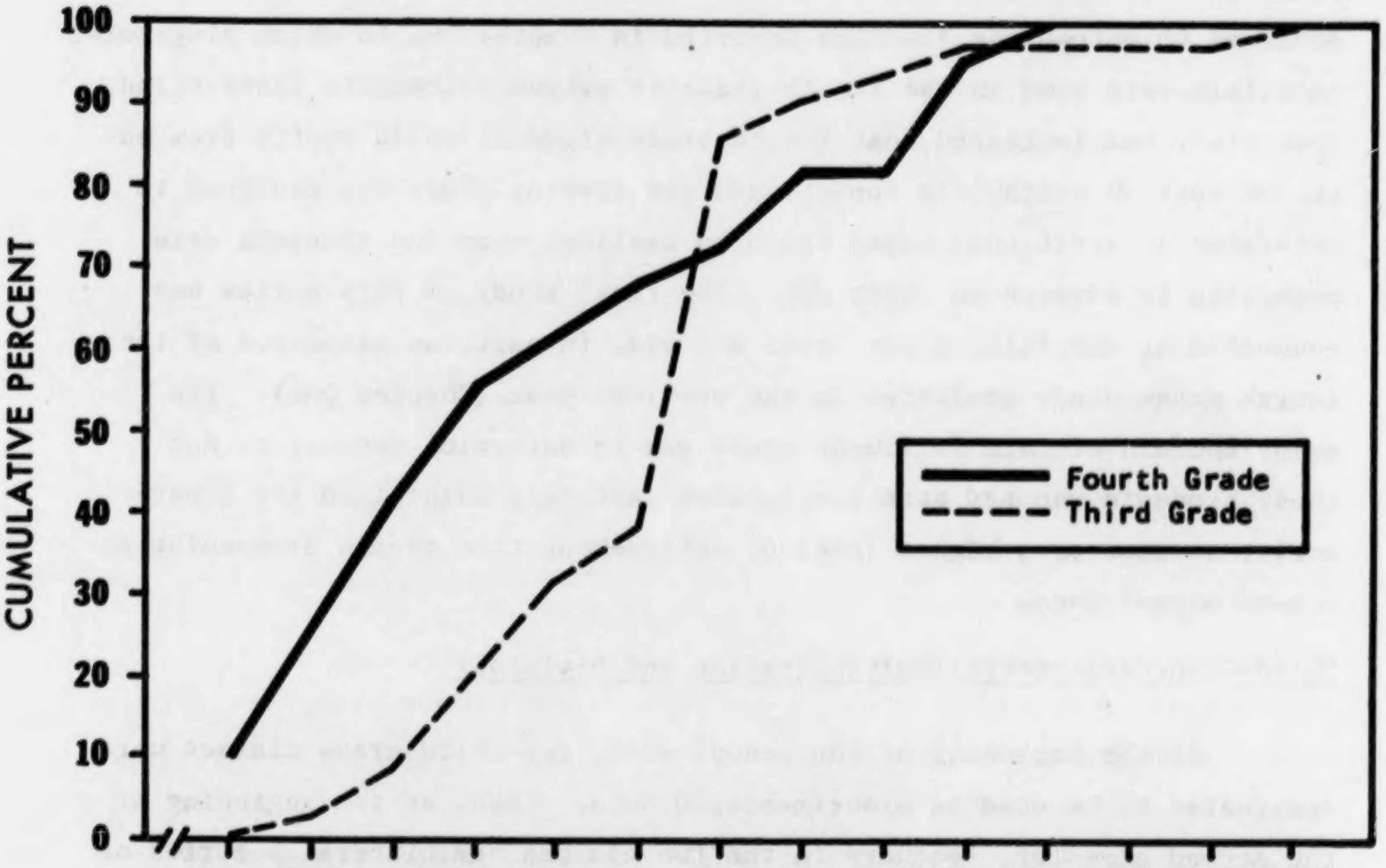

Figure 4.5 Comparison of third- and fourth-grade students in the number of days required to complete spelling program.

levels. In the first study in this series, a selected group of third-grade students received a program in multiplication and division, a topic which is usually considered fourth-grade level work. Although multiplication and division are usually introduced in the third grade, the topic is generally designated as a content area to be developed fully in the fourth grade. Thus, the problem was to determine whether advanced third-grade students could profitably pursue advanced work in this topic. The second study was designed to extend the findings reported in Chapter One in which programmed materials were used in the fourth grade to extend arithmetic instruction. That study had indicated that fourth-grade students could profit from advanced work in arithmetic topics, and the present study was designed to determine if additional gains could be realized when the students were permitted to advance on their own. The final study in this series was conducted at the fifth grade level and was, in part, an extension of the fourth grade study conducted in the previous year (Chapter One). The major concern of this follow-up study was to determine whether or not those students who had been accelerated last year maintained any superiority or reached a higher level of achievement than groups proceeding at a more normal pace.

Third-Grade Arithmetic (Multiplication and Division)

At the beginning of the school year, two third-grade classes were designated to be used as Experimental Groups. Then, at the beginning of the second semester, teachers in the two classes administered a series of diagnostic and achievement tests to all the students in their classes. Those students whose achievement test scores indicated that they could profit from advanced work in multiplication and division were selected from the two classes and served as the Experimental Group (total N=34). Two additional classes, with a total of 56 students, comprised a Control Group. Since the students in the Experimental Group were selected on the basis of high achievement, it was not possible to match the Experimental and Control groups. Table 4.5 presents summary data on IQ and standard

Table 4.5

Summary Data on Standard Intelligence and Achievement Tests Given at the Beginning of the Year to Students in the Third Grade Arithmetic Study

Test	Experimental Group (N=34)		Control Group (N=56)		Group Comparison
	$\overline{X}$	SD	$\overline{X}$	SD	*t*
Otis Quick Scoring Mental Ability Test (IQ)	123.00	11.84	110.84	10.98	4.79***
Metropolitan Achievement Tests: Battery Median Grade Equivalent	4.18	4.68	4.02	6.36	1.38
Metropolitan Achievement Tests: Arithmetic Grade Level	4.01	0.32	3.79	0.45	2.64**
Metropolitan Achievement Tests: Arithmetic, Part A (Concepts)	39.44	3.11	38.30	3.25	1.63
Metropolitan Achievement Tests: Arithmetic, Part B (Computations)	27.91	1.38	27.03	1.63	2.36*

*$p < .05$ (two-tailed test)
**$p < .01$ (two-tailed test)
***$p < .001$ (two-tailed test)

achievement tests given at the beginning of the year. The Experimental Group had a higher IQ (*t*=4.79, *df*=88, *p* <.001), and indicated a higher grade equivalent on the Arithmetic portion of the Metropolitan Achievement Test (*t*=2.64, *df*=88, *p* <.01). This latter difference apparently reflects the difference in performance on the Computation portion of the test where the Experimental Group scored higher (*t*=2.36, *df*=88, *p* <.05) as there was no significant difference between the two groups on the Arithmetic Concepts portion.

Materials. Students in the Experimental Group received conventional instruction from standard textbooks during the first semester and then, during the second semester, were given a programmed textbook in multiplication and division facts. This program contained approximately 2,000 frames divided into 10 teaching units, and is the one used in the previous studies with fourth-grade students. The Control Group continued in its regular third-grade arithmetic textbook throughout the year.

Procedure. At the beginning of the second semester, the students in the Experimental Group were assigned to another room which was under the supervision of one of the regular classroom teachers during their arithmetic class, and work was begun in the multiplication and division program. The general procedure was the same as that specified for the other studies. When the students finished the program, they were again given the battery of diagnostic and achievement tests mentioned above.

In addition to the standard achievement tests, it was desired to have a test with a rather high ceiling so that differences in a high range of achievement would be detected. One of the problems noted in the previous year's studies was that many students obtained near-perfect scores on the posttests, thus making it difficult to determine if any real difference existed between the Experimental and the Control groups at the end of the year. It was decided that the Iowa Test of Basic Skills would be used, but that the standard procedure for administering it would be changed. This test consists of a series of items of increasing difficulty which are grouped for particular grade levels. The students received only that portion which corresponds to their individual academic levels. This procedure was changed so that instead of giving only a portion of the items, all of the items were given which resulted in two tests designated A-1 (arithmetic concepts--136 items) and A-2 (arithmetic problem solving--96 items). Thus, by providing for a wider range of scores, it was felt that perhaps the test would be more sensitive to any differences in the final achievement level of the two groups.

Students in the Control Group continued with the normal third-grade curriculum during the second semester, and were given the battery of diagnostic and achievement tests mentioned above at the end of the year.

Results and Discussion. The pre- and posttest scores in the various tests are presented in Table 4.6. As can be seen from the data, the Experimental Group scored significantly higher than the Control Group on all of the pretests. This is not surprising since the students in the Experimental Group were the most advanced students in two third-grade classes. Moreover, the data also indicated that the Experimental Group maintained its superiority on all of the tests during posttesting (the mean differences in every case being significant beyond the .001 level of confidence). It should be pointed out that on some of the tests, the Control Group showed higher gains from pre- to posttesting than did the Experimental Group. However, this may be primarily an artifact due to the limited range of the particular tests. For example, on the regular Iowa Basic Skills test, the pre- and posttest means for the Experimental Group were 48.62 and 49.97, respectively, indicating a gain of only 1.35 points, while for the Control Group the pre- and posttest means were 40.64 and 45.64, respectively, for a gain of 5.00 points. However, this test contained only 55 items which means that in terms of percentages, the Experimental Groups had an average of 88% correct on the pretest as compared with 74% for the Control Group. Thus, there was less room for improvement for the Experimental Group.

On the other hand, if one examines other tests such as the Multiplication and Division Program test, both groups had fairly low scores on the pretest (means of 12.18 and 9.29 for Experimental and Control Groups, respectively,) and both groups showed fairly substantial gains (i.e., posttest scores of 24.82 and 19.02 for Experimental and Control Groups, respectively). Moreover, on this particular test, the Experimental Group showed a significantly higher gain score than did the Control Group ($\underline{t}$=2.13, $\underline{df}$=88, $\underline{p}$ <.05). Thus, the low ceiling on some of the tests prohibits the making of meaningful comparisons between the two groups in terms of the

Table 4.6

Summary Data on Pre- and Posttests for Students in the Third Grade Arithmetic Study

Test		Experimental Group (N=34)		Control Group (N=56)		Group Comparison
		$\overline{X}$	SD	$\overline{X}$	SD	t
Metropolitan Achievement Test: Computations (47 items)	Pre	29.03	6.62	21.50	4.12	7.70**
	Post	36.68	4.69	32.48	4.59	4.14**
Metropolitan Achievement Test: Concepts (35 items)	Pre	25.41	4.37	16.96	5.65	7.85**
	Post	28.79	4.09	24.55	5.47	4.14**
Iowa Basic Skills: Arithmetic (55 items)	Pre	48.62	4.19	40.64	8.32	5.96**
	Post	49.79	4.10	45.64	6.87	3.70**
M - D Program Test	Pre	12.18	6.89	9.29	5.42	2.06*
(40 items)	Post	24.82	4.14	19.02	5.72	5.50**
M - D Program Final Block Test (40 items)	Post	30.88	4.90	14.04	8.25	12.01**
Modified Iowa Basic Test: A - 1 (Concepts) (136 items)	Post	59.97	11.32	53.52	12.35	1.73
Modified Iowa Basic Test: A - 2 (Problem-Solving) (96 items)	Post	41.35	8.74	34.55	8.96	3.50**

*$p < .05$
**$p < .001$

amount of gain on the achievement tests. The modified Iowa Basic Skills test showed that the Experimental Group scored higher than the Control Group on both the A - 1 (Arithmetic Concepts) and A - 2 (Arithmetic Problem-Solving) sections (see Table 4.6), but the groups differed significantly only on the A - 2 section of the test (t=3.50, df=88, $p<.001$). Thus, the results reflect the same pattern noted in the Metropolitan Achievement pretest scores--that is, the Experimental Group was significantly higher on the computational portion of the test, but not on the arithmetic concepts portion. Thus, the results would seem to permit the implication that the Experimental Group did maintain its superiority in overall arithmetic achievement level during the course of the study.

It is also of interest to again note the individual differences in rate of learning as reflected by the number of days required to complete the program (see Figure 4.6). All of the students were able to finish the program within the allotted time and required an average of 40.21 days to completion (SD=11.30). As in the other studies, there were again wide differences in the amount of time required to complete the program. As the figure indicates, four of the students were able to complete the program in 16-20 days, while the slowest eight students required 51-55 days. Thus, even with a homogeneous, above-average group like that in the present study, one still finds impressive individual differences in rate of learning.

In summary, then, the use of programmed materials permitted high-achieving third-grade students to pursue advanced work in arithmetic, and according to the standardized achievement tests, to maintain their superiority over students receiving a full year of conventional classroom instruction. Moreover, all of the students were able to finish the program within the allotted time, and probably could have done additional work if appropriate materials had been provided.

Fourth Grade Arithmetic (Multiplication and Division and Fractions)

In the work reported in Chapter One, it was found that a fourth grade class receiving programmed instruction was able to go into advanced

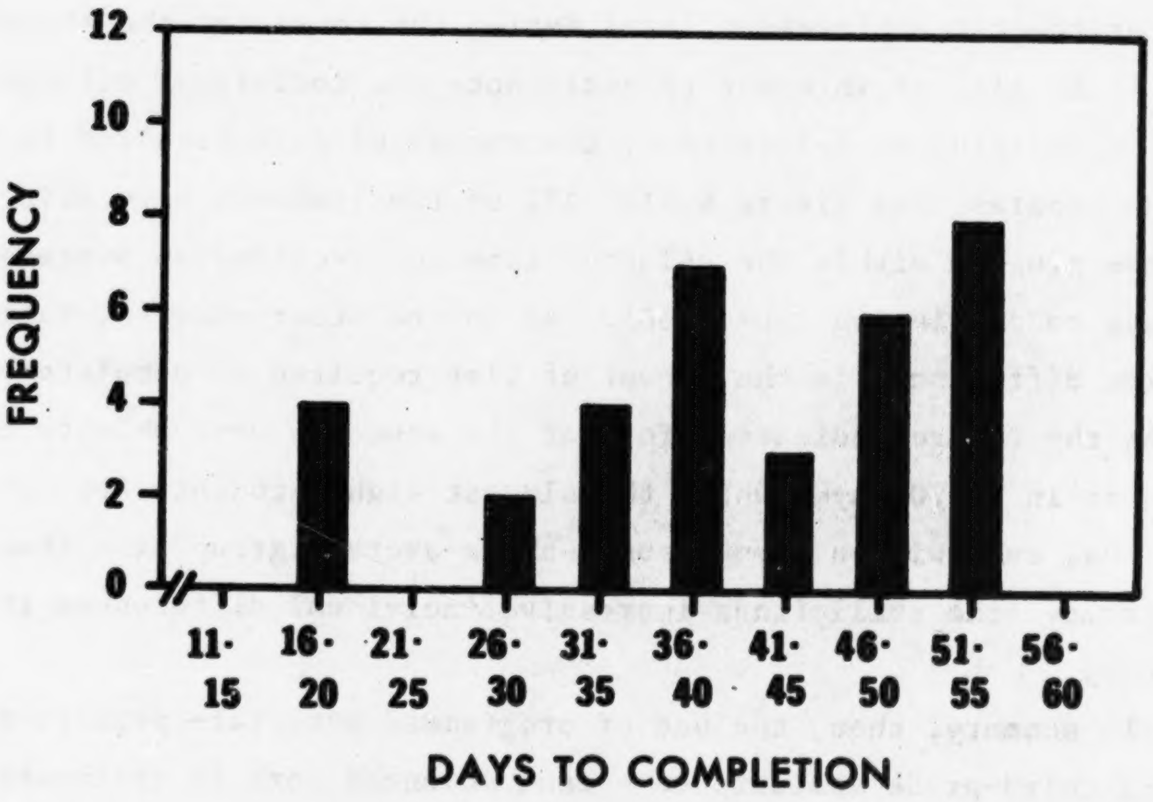

Figure 4.6 Number of days to complete multiplication and division program by third-grade students (N=34).

work in fractions with no loss in learning or retention of the usual fourth-grade arithmetic curriculum. These groups did as well as other groups on tests covering the regular curriculum, and showed a significantly higher level of performance, as might be expected, on an achievement test in fractions. These results suggested that it might be possible to extend arithmetic instruction even more at this grade level; thus, the objective of the present study was to see just how far students could progress when given the opportunity and the proper materials.

A group of 48 students from two fourth-grade classes was selected to receive individualized instruction in multiplication and division and fractions. These students were to spend the entire year in individualized arithmetic instruction and were to progress as far as possible in the series of arithmetic materials. Since the purpose of the present study was simply to determine the effectiveness of providing opportunities for extended learning, there was no Control Group as such in the study. However, it was felt that some comparisons with a conventionally-taught class might be useful, so some test data was obtained from two additional fourth-grade classes (N=50) at the end of the year. In addition, scores were available on IQ and the Metropolitan Achievement Test given at the beginning of the year, and these data are summarized in Table 4.7. There were no reliable differences between the two groups in IQ or in Metropolitan Battery Median, but the Experimental Group did have a higher Average Arithmetic Grade Equivalent ($\underline{t}$=2.06, $\underline{df}$=96, $\underline{p} < .05$), which resulted apparently from the higher score on the Arithmetic Computation Subtest ($\underline{t}$=2.65, $\underline{df}$=96, $\underline{p} < .01$).

Materials. The programmed textbook in multiplication and division facts described in the third grade study was again used. When students finished the program, they began supplementary work with advanced multiplication and division materials prepared by the Center staff and Baldwin-Whitehall teachers. Four units of supplementary work were developed from conventional materials with pre- and posttests for each unit. When the students finished the supplementary material, they received a programmed textbook in fractions, which consisted of approximately 1,700 frames divided into 12 teaching units. As the study progressed, it was found that many of the students were completing the fractions program in addition to

Table 4.7

Summary Data on Standard Intelligence and Achievement Tests Given at the Beginning of the Year to Students in the Fourth-Grade Arithmetic Study

Test	Experimental Group (N=48)		Control Group (N=50)		Group Comparison
	$\overline{X}$	SD	$\overline{X}$	SD	t
Otis Quick Scoring Mental Ability Test (IQ)	110.67	12.50	111.36	13.08	0.27
Metropolitan Achievement Tests: Metropolitan Battery Median	4.72	0.81	4.66	1.32	0.27
Metropolitan Achievement Tests: Average Arith. Grade Equivalent	4.79	0.71	4.36	1.27	2.06*
Metropolitan Achievement Tests: Arithmetic Computation	34.96	6.18	30.64	9.50	2.65**
Metropolitan Achievement Tests: Arithmetic Problem-Solving	23.21	6.26	22.16	8.08	0.71

*$p < .05$ (two-tailed test)
**$p < .01$ (two-tailed test)

all of the multiplication and division work; thus, it was decided that they would be permitted to enter into the supplementary fractions work (five units) which was developed for the fifth grade and which will be described in the next study. Thus, there were four basic sets of materials which were used by these students: (1) a programmed textbook in multiplication and division facts, (2) supplementary materials in multiplication and division, (3) a programmed textbook in fractions, and (4) supplementary materials in fractions. The Control classes received conventional classroom instruction throughout the year with no special instructions being given to the teachers.

Procedure. The same procedure followed in the other studies was again used. Students in the Experimental Group worked on the programmed materials during their regular arithmetic period each day and were permitted to advance through the material as rapidly as they indicated mastery of each unit. Detailed records were kept concerning performance and amount of time spent in each unit, and a number of achievement tests were given at the beginning and end of the year to assess the amount of gain in achievement level.

Results and Discussion. Summary information on the pre- and post-testing for the Experimental Group is presented in Table 4.8, and indicates that there were gains on all of the tests (all gains were statistically significant at $p < .001$). Unfortunately, data on these particular tests were not available for a Control Group so that group comparisons could be made. However, data were obtained on the modified Iowa Basic Skills Test (i.e., the wide-range test described in the third grade arithmetic study) at the end of the year for both groups so that a comparison could be made regarding end-of-year performance level. These data are presented in Table 4.9 and indicate that the two groups did not differ significantly on either portion of the test. The two groups had almost identical scores on the A - 1 (Arithmetic Concepts) section with means of 69.83 and 70.46 for the Experimental and Control Groups, respectively. The Experimental Group did score higher than the Control on the Arithmetic Computation section (means of 49.96 and 47.24, respectively), but the difference was not large enough to be statistically significant (t=1.02, df=96, $p > .05$). Thus, these data, along with those from the Metropolitan Achievement Tests mentioned earlier, seem to indicate that while the students receiving programmed instruction did not reach a higher achievement level at the end of the year, they did reach an achievement level equal to that of the conventionally taught classes. A similar result was found in the fourth grade study of the previous year reported in Chapter One; that is, a group of students given advanced work in fractions (the same program used in the present study) demonstrated a general achievement level at the end of the year equal to that of groups who had not received such

Table 4.8

Summary Data on Pre- and Posttests for Fourth-Grade Students Receiving Individualized Arithmetic Instruction

Test	Pretest		Posttest		Gain	
	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
Inventory Test (28 items)	24.71	4.35	26.24	3.51	1.53	2.13
M - D Program Test (40 items)	18.35	7.80	27.98	6.15	9.63	6.28
M - D Supplement Test (25 items)	11.41	5.12	19.17	5.30	7.76	4.29
Fractions Program Test (50 items)	23.29	9.17	39.98	9.74	16.69	7.40
Fractions Supplement Test (20 items)	3.69	2.57	13.98	5.81	10.29	4.73

advanced material, but who had spent the entire year in the regular fourth grade curriculum. In addition, this group scored significantly higher than the others on an achievement test in fractions, as might be expected. Hence, those students were seen to profit from the new material, with no loss in achievement in the regular arithmetic curriculum.

It may be informative to compare the performance of the present group with that of the groups in last year's study. However, these comparisons must be regarded as rather tenuous since it is not possible to specify how similarly the groups were treated, how well-matched they were, etc. The summary data on the fractions program posttest for the groups in last year's study and this year's study are presented in Table 4.10.

Table 4.9

Comparison of Fourth-Grade Experimental and Control Groups on the Modified Iowa Basic Arithmetic Test Given at the End of the Year

Test	Experimental Group (N=48)		Control Group (N=50)		Group Comparison
	$\bar{X}$	SD	$\bar{X}$	SD	t
A - 1: Arithmetic Concepts (136 items)	69.83	17.88	70.46	13.62	-0.19
A - 2: Arithmetic Problem-Solving (96 items)	49.96	14.45	47.24	11.87	1.02

Table 4.10

Comparison of Fourth-Grade Classes During 1962-63 and 1963-64 Studies on Fractions Program Posttest

Group	N	Mean	SD
Program (1963-64)	48	39.98	9.74
Program (1962-63)	52	43.75	4.85
Control (1962-63)	59	34.48	9.92

The comparison of last year's program with the present group indicated that the present group did not do as well on the test ($\underline{t}=2.42$, $\underline{df}=98$, $\underline{p}<.05$). However, the present Experimental Group did do significantly better than the Control Group in the earlier study ($\underline{t}=2.88$, $\underline{df}=105$, $\underline{p}<.01$). Again, however, it must be pointed out that these comparisons can only be regarded as tentative. While the data do suggest that the current group did realize some additional achievement over a conventionally-taught class, this conclusion must be viewed with reservation.

The main objective of the present study, of course, was to obtain information about the acceleration of arithmetic instruction. Thus, interest was focused on the amount of material covered by the students in the Experimental Group. Approximately 96% of the students successfully completed all of the material through the fractions program, and 88% of the students completed all of the work including the supplementary fractions work. Thus, the students were able to complete all of their fourth grade arithmetic work, plus a substantial amount of advanced work of approximate fifth grade difficulty. In order to gain more information about individual differences in rate of learning, distributions were obtained for the number of days required to finish various amounts of the material, and these data are presented in Figure 4.7. This figure is cumulative in nature and is read as follows:

(1) the top panel (panel D) shows the distribution of students finishing all four sets of material in multiplication and division, and fractions in different periods of time;

(2) the second panel (panel C) gives the distribution of students completing all of the multiplication and division work plus the fractions program;

(3) the next panel (panel B) gives the distribution of students completing only the multiplication and division program and supplementary work;

(4) the bottom panel (panel A) gives the distribution of students finishing only the multiplication and division program.

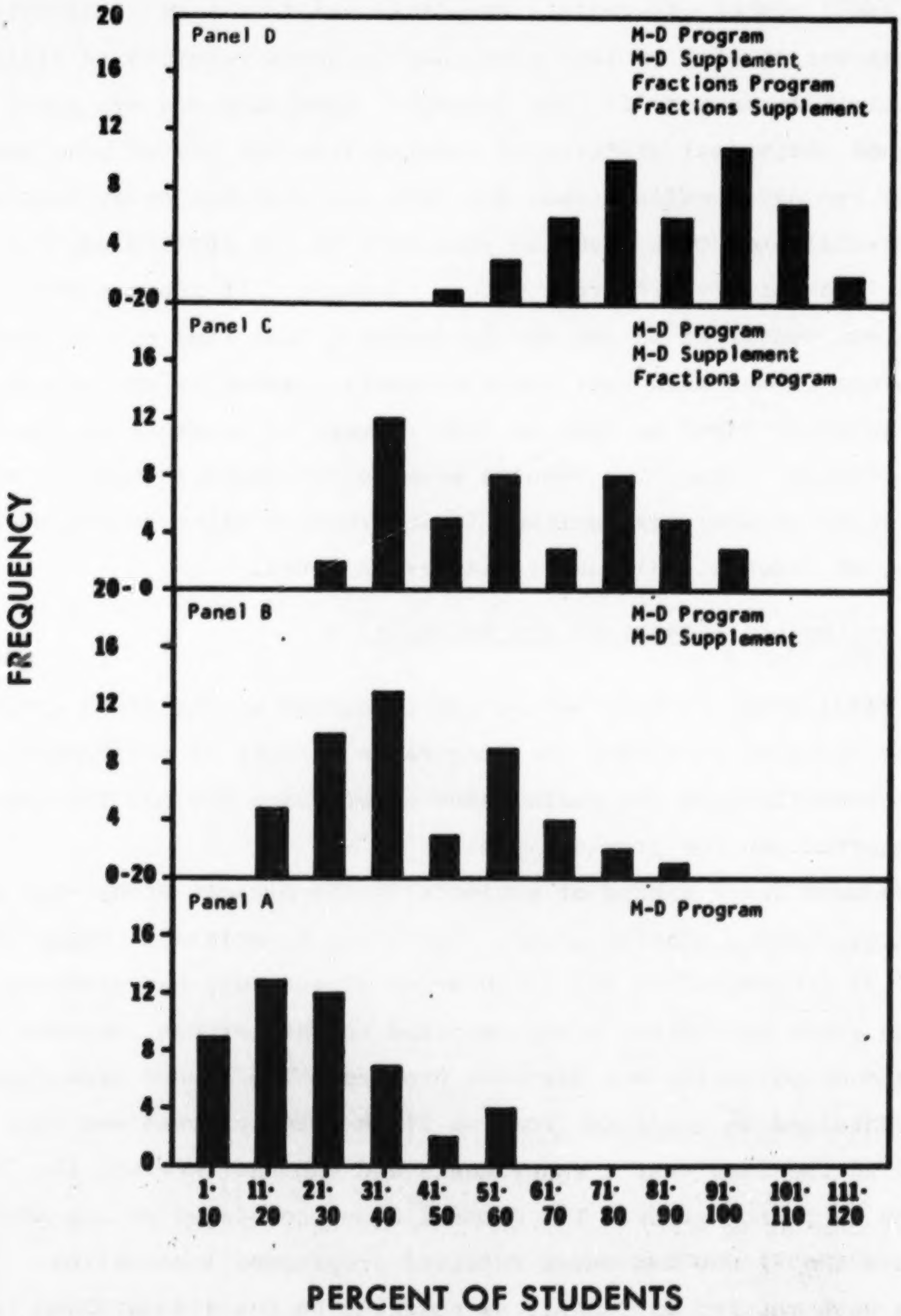

Figure 4.7 Number of days to complete various amounts of arithmetic material by fourth-grade students (N=48).

The great amount of overlap is readily apparent. For example, it can be seen that a small number of students completed all four sets of materials in the same amount of time or less than some students required to finish only the multiplication and division program. Thus, one can see again the accentuation of individual differences arising from the use of programmed materials. These data indicate that not only can students cover considerably more material than that normally presented in the curriculum, but that they do it at widely different rates. Moreover, it appears that these students are not penalized in any way by spending less time on the regular curriculum since it was seen that these students reached an end-of-year achievement level at least as high as that reached by students in the conventional classroom. Thus, the results seem to corroborate those from last year's study showing that arithmetic instruction might be extended considerably, at least at this particular grade level.

Fifth Grade Arithmetic (Fractions and Decimals)

The final study in this series was conducted at the fifth grade level and was designed to assess the long-range effects of programmed instruction by investigating the performance of students who had received programmed instruction the previous year.

There were three groups of subjects in the present study--two experimental groups and a control group. The first Experimental Group (E_1) consisted of 74 students from two fifth-grade classes who had participated in the fourth grade arithmetic study reported in Chapter One, and who had received the multiplication and division program. The second Experimental Group (E_2) contained 64 students from two fifth-grade classes who also had participated in the last year's experiments and who had received the fractions program in fourth grade. The Control Group consisted of two additional classes (N=57) who had never received programmed instruction. The three groups were matched as closely as possible on the standardized IQ and achievement tests mentioned in the earlier studies, and summary data on these tests is presented in Table 4.11. As the data indicate, the groups were well-matched as there were no reliable differences among the groups on any of the tests.

Table 4.11

Summary Data on Standard Intelligence and Achievement Tests Given at the Beginning of the Year to Students in the Fifth Grade Arithmetic Study

Test	Experimental-E_1 (N=74)		Experimental-E_2 (N=64)		Control (N=57)		Group Comparison
	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\underline{F}$
Otis Quick Scoring Mental Ability Tests (IQ)	112.85	11.15	112.30	9.82	114.4	10.71	<1.00
Stanford Achievement Tests: Stanford Battery Median	6.20	1.08	6.22	0.96	5.97	0.84	1.22
Stanford Achievement Tests: Average Arithmetic Grade Equivalent	6.42	1.07	6.46	0.95	6.17	0.93	1.51
Stanford Achievement Tests: Arithmetic Reasoning Grade Equivalent	6.50	1.12	6.47	1.05	6.22	1.03	1.26
Stanford Achievement Tests: Arithmetic Computation Grade Equivalent	6.32	1.18	6.45	1.09	6.08	1.09	1.66

Materials. The basic materials used by the Experimental Groups in this study were a programmed textbook in fractions, supplementary fractions work developed by teachers, and a programmed textbook in decimals. As mentioned in the previous study, the fractions program contains approximately 1,700 frames divided into 12 teaching units. The supplementary materials (5 units) were designed to extend the coverage of the fractions program, and were developed by the fifth-grade teachers utilizing the techniques described earlier. The decimals program is designed to follow the fractions program and contains approximately 2,700 frames divided into 13 teaching units. No supplementary lessons in decimals were developed.

The Control classes received conventional classroom instruction using the regular fifth-grade arithmetic textbooks and other materials which the teacher wished to use.

Procedure. Students in Experimental Group E_1 were given the fractions program at the beginning of the year and began work in it immediately. As they completed the program, they advanced to the supplementary fractions work and then to the decimals program. The second Experimental Group, E_2, had already had the fractions program in fourth grade and so received classroom instruction in advanced fractions during the first semester. At the beginning of the second semester, this group was then entered into the decimals program. Both Experimental Groups, upon completion of the decimals program, returned to regular classroom instruction and began work in modern mathematics using a conventional textbook. During the program work, the general procedure outlined previously was followed, allowing the students to progress as rapidly as possible through the material. (The decimals program did not contain ready-made unit pre- and posttests so these were developed by the teachers in collaboration with Center Staff members.) The Control Group received conventional classroom instruction throughout the year following the normal fifth grade arithmetic curriculum.

Results and Discussion. All groups were given a battery of achievement tests at the beginning of the year and at the end of the year, and these data are presented in Table 4.12. The groups did not differ significantly on any of the pretests except the fractions program test (F=5.54, df=2/192, $p < .01$), and the fractions supplement test (F=3.68, df=2/192, $p < .05$). On the first test, the E_2 Group scored higher than either of the other two groups, as might be expected since this is the group which received the fractions program last year. The data on the fractions supplement test also indicated that the E_2 Group again had the highest mean number correct, although the differences among groups were not as large as on the fractions program test. Thus, these data indicate that the group which received the fractions program last year and which exhibited a higher achievement level at the end of fourth grade still maintained its higher achievement level at the beginning of the study.

The posttest data also indicated highly significant differences among the three groups on the fractions program test (F=22.17, df=2/192, $p < .001$), and on the fractions supplement test (F=59.22, df=2/192, $p < .001$). However, there was some difference in the order of the means from the pretest data. Individual t-tests were computed to compare the various groups and indicated that on the fractions program test, the E_1 Group scored significantly higher than the E_2 Group (t=3.12, df=126, $p < .01$) and the Control Group (t=6.35, df=129, $p < .001$), and that the E_2 Group scored significantly higher than the Control Group (t=3.31, df=119, $p < .01$). The fact that the E_2 Group still scored higher than the Control Group on the fractions program test at the end of the year lends some support to the finding that the extended arithmetic instruction of last year produced a stable higher achievement level in fractions work. The superiority of the E_1 Group over the E_2 Group may be due primarily to the fact that the E_1 Group had been exposed to the fractions program more recently than had the E_2 Group and that, hence, their retention of the material had not shown the decline characteristically found when retention is measured after some fairly long period of time. On the fractions supplement test, the E_1 Group again scored significantly higher than both the E_2 Group (t=10.57, df=136, $p < .001$), and the Control Group (t=2.15, df=129, $p < .05$). However, the E_2 Group, in contrast to the results found on the fractions program test, was significantly lower than the Control Group (t=7.85, df=119, $p < .001$). The reason for this discrepancy in results is not clear since

Table 4.12

Summary Data on Pre- and Posttests for Students in the Fifth Grade Arithmetic Study

Test		Experimental-E_1 (N=74)		Experimental-E_2 (N=64)		Control (N=57)		Group Comparison
		$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	F
Iowa Basic Skills: Arithmetic	Pre	44.22	11.33	39.94	11.51	43.32	11.27	2.84
(71 items)	Post	48.69	11.50	47.97	10.68	47.47	9.85	<1.00
Inventory Test	Pre	20.66	4.79	19.31	5.43	21.05	5.22	<1.00
(30 items)	Post	21.50	5.19	22.78	4.89	21.89	4.63	<1.00
M - D Program Test	Pre	32.43	5.55	31.59	6.01	33.00	5.66	<1.00
(40 items)	Post	33.46	5.86	34.61	5.57	35.58	4.67	2.64
Fractions Program Test	Pre	33.36	10.62	37.23	9.78	30.82	10.95	5.54**
(50 items)	Post	44.04	6.92	39.86	8.48	34.25	9.84	22.17***
Fractions Supplement Test	Pre	5.19	3.21	5.30	2.02	4.04	3.43	3.68*
(20 items)	Post	14.18	3.68	7.45	3.71	12.77	3.67	59.22***
Modified Iowa Basic Test A - 1 (136 items)	Post	83.81	17.43	83.59	18.39	85.04	15.13	<1.00
Modified Iowa Basic Test A - 2 (96 items)	Post	58.28	14.84	60.22	15.78	56.61	14.09	<1.00
Decimals Program Test	Pre	8.66	5.83	4.91	5.57	(Not given to Control Group)		t=3.86***
(35 items)	Post	22.89	7.42	20.81	7.87			t=1.59

*p < .05
**p < .01
***p < .001

the E_2 Group did have a semester of teacher instruction in advanced fractions at the beginning of the year, and should have learned the topics covered in the fractions supplement test. However, it might again reflect some "recency" effects--that is, the Control Group, since it did not go into decimals work, had spent most of its time in fractions and thus had probably had exposure to the advanced material covered by the test more recently than had the E_2 Group. These data also seem to indicate that the supplementary units developed by the teachers were quite effective teaching instruments since the E_1 Group did score significantly higher on the test than the Control Group although the E_1 Group had not been exposed to the fractions material as recently as the Control Group. None of the other posttest results nor the scores on the Modified Iowa Basic Skills Test indicated any reliable differences among the three groups.

The decimals program test was given only to the E_1 and the E_2 Groups and indicated that on the pretest, the E_1 Group scored significantly higher than Group E_2, ($\underline{t}$=3.86, $\underline{df}$=136, $\underline{p}<.001$) but that on the posttest, the two groups did not differ significantly ($\underline{t}$=1.59, $\underline{df}$=1.36, $\underline{p}>.05$). Thus, the two groups reached equivalent achievement levels as a result of the programmed course in decimals.

Figure 4.8 presents the distribution of students in Group E_1 finishing different amounts of the programmed work (since students in the E_2 Group worked only in the decimals program, their data is not included in the graph). All of the students in Group E_1 finished the fractions program, 95% finished the fractions supplement, and 80% finished the decimals program. Figure 4.8 is a composite graph like the one in the fourth grade study showing the variations in amount of time required to complete various portions of the material, and once again reflects the wide variations in student performance as measured by time to completion.

Thus, these data again indicate that the use of programmed materials enabled students to cover more material than they normally do, and that these students reached an end-of-year achievement level which was generally higher than that of a conventionally-taught class. Moreover, the performance of the E_2 Group suggests that the beneficial effects

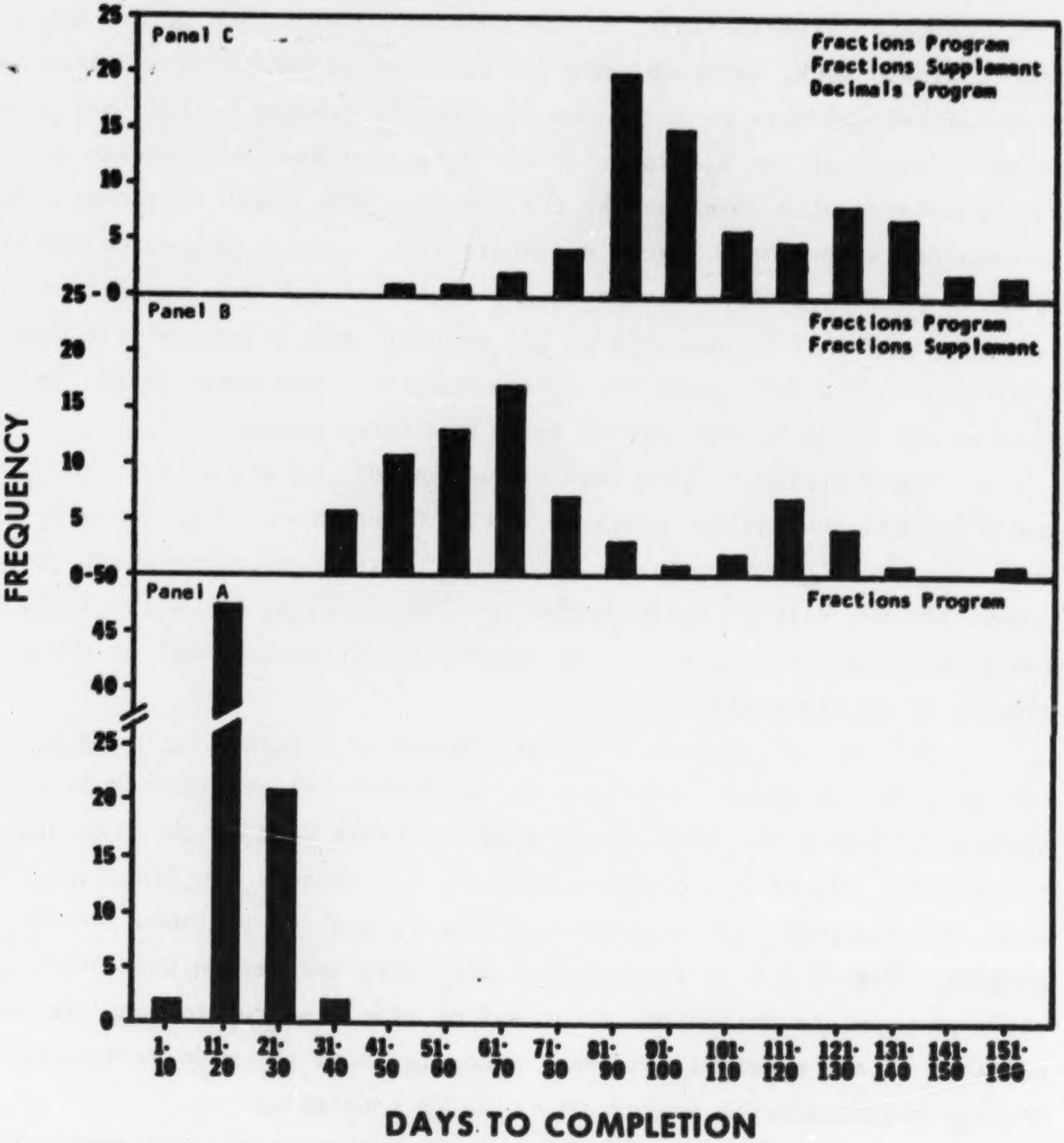

Figure 4.8 Number of days to complete various amounts of arithmetic material by fifth-grade students (N=74).

accruing from programmed instruction are stable and long lasting, and that students who do receive advanced material in a particular grade continue to show some superiority over students proceeding at a more normal pace.

Summary of the Arithmetic Studies

In summary, these studies indicate that programmed materials are generally effective in individualizing and extending arithmetic instruction. In all of the studies, the groups using programs were able to cover considerably more material than similar groups receiving conventional in-instruction, and moreover, the achievement level of the Experimental Groups at the end of the year was, in almost all cases, equal to or above that of the Control Groups. Thus, these results, in conjunction with those from the spelling studies discussed earlier, lend support to the use of programs as an effective means of providing individualized, extended instruction in the elementary grades.

Individualized Use of Programmed Instruction as a Review Tool

In addition to the general use of programs as instructional tools, it would appear that programmed instruction would also have some merit when used as a review tool. In the typical classroom situation, a good deal of time at the beginning of the school year is devoted to the review of materials covered in the previous year's work. In general, such review is helpful since there are students who undoubtedly have forgotten part of the materials which they had previously mastered. However, it is also true that many students may not need any review or may need review only on selected topics. Thus, a general review for all students may be of questionable value, and certainly consumes time which could be spent on new work. Consequently, it was decided to use programmed instruction as a review tool, and in keeping with the general objectives of the research program, to individualize the review period as much as possible.

Six fifth-grade classes which had had no previous experience with programmed instruction were selected and given various kinds of review of multiplication and division facts which had been taught in fourth grade. Two classes received a multiplication and division program (Program Group, N=52), two classes used workbooks as review tools (Workbook Group, N=45), and two classes received review as a part of the conventional classroom procedure, (Control Group, N=49). An attempt was made to match the groups as closely as possible on intelligence and achievement level and Table 4.13 gives descriptive information for the various groups on the different tests. As the table indicates, the matching was effective with the exception that the Control Group had a higher mean IQ than the Program Group ($t=2.54$, $df=99$, $p < .05$). None of the groups differed significantly on any of the other measures.

Materials. The students in the Program Group were given a programmed textbook in multiplication and division facts which was used in the third and fourth grade studies. The Workbook Group received a typical fifth-grade workbook with which to review multiplication and division skills, and lessons were developed from this material and presented in the daily arithmetic sessions. The Control Group used its regular fifth-grade textbook in their review work in conjunction with regular teacher instruction. In addition, teachers in this group were allowed to use any other materials which they would ordinarily use in reviewing such material.

Procedure. Utilizing the individualization techniques described previously, students in the Program Group reviewed only those topics with which they were having difficulty; that is, they would work only on those portions of the program which their pretest performance indicated they did not know. Thus, the student was allowed to progress through the material at his own pace and was not required to spend time on material which he had already mastered. The program work was completed during a part of the regular arithmetic period with the teacher available to give individual attention to those students having particular difficulties. When the students completed their program work, they were entered into classwork on new arithmetic material.

Table 4.13

Summary Data on Standard Intelligence and Achievement Tests Given at the Beginning of the Year to Students in the Fifth Grade Individualized Review Study

	Program Group (N=52)		Workbook Group (N=45)		Control Group (N=49)		Group Comparison
	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\underline{F}$
Otis Quick Scoring Mental Ability Tests (IQ)	**104.96**	**12.68**	**108.98**	**11.64**	**111.08**	**11.38**	**3.45***
Stanford Achievement Tests: Stanford Battery Median	**5.47**	**0.96**	**5.28**	**1.16**	**5.68**	**1.09**	**1.66**
Stanford Achievement Tests: Arithmetic Grade Equivalent	**5.73**	**0.94**	**5.36**	**1.16**	**5.75**	**1.00**	**2.12**
Stanford Achievement Tests: Arithmetic Reasoning Grade Equivalent	**5.92**	**0.97**	**5.49**	**1.34**	**5.96**	**1.20**	**2.31**
Stanford Achievement Tests: Arithmetic Computation Grade Equivalent	**5.55**	**1.03**	**5.28**	**1.19**	**5.49**	**1.03**	**1.00**

***$p < .05$ (df=2/143)**

Students in the Workbook Group were given daily assignments in the Workbook which they also completed during part of the regular arithmetic period. To some extent, these students were allowed to work at their own pace, but in general, were held to the pace of the class as a whole since they also received some conventional teaching during the arithmetic period. Unlike the Program Group, they were not allowed to skip through the material, but covered all of the topics in the review.

Students in the Control Group proceeded with regular fifth-grade work without a formal period of review. No attempt was made to individualize the review work, but rather, the teachers integrated the review into their regular classroom work, adhering to their usual teaching methods.

The termination date for the study was determined by ascertaining the amount of material covered by students in the Program Group and using this amount of material as the criteria for the other groups. Thus, when the other groups had covered the same amount of material as that in the multiplication and division program, the review period was ended.

Results and Discussion. All students in the study were given a series of pretests in arithmetic at the beginning of the study, and then received the same tests as posttests at the end of the review period. (For students in the Program Group, the posttests were administered as each student finished the program; for the other two groups, the posttests were administered to all students at the end of the review period.) Summary data for each group are presented in Table 4.14 and indicate that, in general, the three groups showed generally equivalent performance during pretesting. The Control Group did have a higher mean score than the other two groups on the Inventory Test ($F=5.22$, $df=2/143$, $p<.01$), but the groups showed equal performance on the remaining pretests.

During posttesting, the three groups did not differ significantly on the Inventory Test or on the Iowa Basic Skills Test. However, if one examines the gain scores from pre- to posttesting, it can be seen that on both tests, the Program and Workbook Groups showed significantly more gain than did the Control Group. In addition, the Program Group showed a higher gain than the Workbook Group on the Iowa Basic Skills Test ($t=3.05$, $df=95$,

$p < .01$). On the multiplication and division program posttest, both the Program Group and the Workbook Group scored higher than the Control Group (F=5.18, df=2/143, $p < .01$), but did not differ significantly from each other. Finally, all the groups were given the multiplication and division program final block test (see Table 4.14), and the Program Group performance on this test was much higher than that of either of the other two groups (F=19.80, df=2/143, $p < .001$). One might, of course, expect the Program Group to do better on such a test because the test was specific to the program. While it is true that the test did cover the material taught in the program, the content of the test represented the material which was covered by all groups. Thus, even though the test was developed for the program, it should have been respresentative of the material which all groups were supposed to review. These results indicate that the use of programmed materials was effective as a review tool. In all cases, the Program Group showed more improvement than the Control Group although the differences in performance were not always necessarily large.

Perhaps the most interesting aspect of the study is the difference in the amount of time it took each group to complete the review work. It will be recalled that students using programmed materials were permitted to finish the review work as quickly as they could. Then, using the amount of material covered by the Program Group as a criterion (i.e., the content of the review material), the other two groups were allowed to continue working until they had reached the same criterion. Students in the Program Group required a mean of 21.33 days (SD=3.00) to complete the material, while in the other two groups, all students worked 40 days to complete the same amount of material. Thus, the Program Group was able to complete the material in almost half the time required by more conventional teaching methods. Equally interesting is the range of time required by students in the Program Group. Five students completed the material in only 17 days, while the slowest eight students required only 26 days--still a rather impressive saving of time compared to the other groups.

Table 4.14

Summary Data on Pre- and Posttests for Students in the Fifth Grade Individualized Review Study

Test		Program Group (N=52)		Workbook Group (N=45)		Control Group (N=49)		Group Comparison
		$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\underline{F}$
Iowa Basic Skills: Arithmetic	Pre	33.10	9.87	34.89	13.93	36.65	9.19	1.30
(71 items)	Post	40.54	10.04	38.53	14.50	37.98	9.87	<1.00
Inventory Test	Pre	15.42	5.37	14.58	6.09	18.20	5.83	5.22*
(30 items)	Post	19.25	5.40	17.64	6.96	19.45	8.24	<1.00
M - D Program Test	Pre	27.85	5.89	27.02	8.24	29.02	7.48	<1.00
(40 items)	Post	30.85	5.61	31.96	6.77	27.69	7.66	5.18*
M - D Program Final Block Test (40 items)	Post	34.37	3.96	26.92	9.40	25.12	9.22	19.80**

*p<.01 (df=2,143)
**p<.001 (df=2,143)

In summary, these data indicate that a defined review period with individual assignments was more effective than the conventional method of integrating review into the regular classroom work. While the Program and Workbook Groups were generally equivalent on the posttest measures, the students using workbooks were required to work a full 40 days as compared with an average of 21.33 days for those students using the program. Thus, considering both the test data and the amount of time worked, the use of programmed instruction appears to be a more effective means of review than the other two methods.

Summary and Conclusions

This report has discussed a series of studies in which programmed instruction was used as a means of individualizing and extending instruction in a public school system. Previous research (Chapter 1) had indicated that programmed materials could be used as a means of implementing instruction in the intact classroom. The work reported in this chapter was directed toward determining to what extent such materials could be used to provide a more individualized course of instruction for each student. A series of five pilot studies was conducted at the third, fourth, and fifth grade levels using commercially available programs in spelling and in various arithmetic topics. An additional study was also carried out at the fifth grade level to evaluate the effectiveness of programmed instruction as a means for providing individualized review of material covered the previous year.

The technique of individualization which was developed involved breaking each instructional topic down into a number of small teaching units with tests to cover the material within that unit. Then, before beginning work in a unit, each student was given the unit test as a pretest and if his score were above a predetermined criterion, he would be allowed to skip that unit and go on to the next unit where he was again given the pretest. In this way, the student did not have to spend time going over material which he already knew, but was able to go into new material, and was able to advance at his own pace. In addition, supplementary materials were developed in many areas which permitted the student to pursue advanced work on an individualized basis after he had finished the programmed work. In almost all of the studies, comparable groups of students receiving conventional instruction were used as control groups so that comparisons could be made in regard to amount of material covered, amount of time spent in various topics, and final achievement level as measured by a number of different tests.

The general conclusions warranted by the present data are the following:

1. Programmed instruction appears to be an effective means of providing individualization of instruction.

2. By breaking the material down into small instruction units, it can be seen that many students know a certain amount of material before they begin receiving instruction.

3. Individual students, even within homogeneously-grouped classes, vary a great deal in the rate at which tehy can advance through a given set of instructional materials, and in the amount of material which they can master within a given time period.

4. The individualized use of programmed instructional materials enabled students to cover a given amount of material in less time than than required by students receiving conventional instruction.

5. Because students using programmed materials were able to advance through the curriculum more rapidly than usual, they were able to enter into advanced work in a given topic which was not normally available to them at their particular grade level.

6. In almost all cases, the achievement level of students receiving programmed instruction was equal to or greater than that of students receiving conventional instruction as measured by standardized achievement tests.

7. In addition to its effectiveness as a teaching tool, programmed instruction appears to be an effective method of presenting review.

It would appear, then, that the present data, coupled with that from previous studies, strongly indicate that well-designed programs can be quite effective in implementing and supplementing current instructional practices in the public schools. Moreover, the variation in the level of entering behavior of students and the rate of progress ghrough given instructional materials indicate that there is an immediate need for the development of instructional practices which take into account (1) the knowledge and skills brought into the learning situation by the learner, and (2) an adaptation of the instructional program to the needs of the student. As indicated earlier, such instructional practices involve the development of procedures for individualizing instruction. The limited data presented in this paper suggest that programmed instruction may be a basis from which to begin development of the procedures for this kind of individualization.

References

Bottel, Morton. Multi-level speller. Penns Valley Publishers, Inc., State College, Pa., 1959.

Greene, Harry A. The new Iowa spelling scale. Iowa City, Iowa: The State University of Iowa Press, 178 p.p.